PTO/SB/05 (08-00)
Approved for use through 10/31/2002 OMB 0651-0032
U S Patent and Trademark Office; U.S DEPARTMENT OF COMMERCE

Please type a plus sign (+) inside this box → [X]

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Attorney Docket No.	CL000841
First Inventor	LUDUNGA et al
Title	ISOLATED HUMAN SECRETED ....
Express Mail Label No.	

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

ADDRESS TO: Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

1. [XX] Fee Transmittal Form (e.g., PTO/SB/17) *(Submit an original and a duplicate for fee processing)*
2. [ ] Applicant claims small entity status. See 37 CFR 1.27.
3. [XX] Specification [*Total Pages* 63] *(preferred arrangement set forth below)*
   - Descriptive title of the invention
   - Cross Reference to Related Applications
   - Statement Regarding Fed sponsored R & D
   - Reference to sequence listing, a table, or a computer program listing appendix
   - Background of the Invention
   - Brief Summary of the Invention
   - Brief Description of the Drawings *(if filed)*
   - Detailed Description
   - Claim(s)
   - Abstract of the Disclosure
4. [XX] Drawing(s) *(35 U.S.C. 113)* [*Total Sheets* 43]
5. Oath or Declaration [*Total Pages*]
   a. [ ] Newly executed (original or copy)
   b. [ ] Copy from a prior application (37 CFR 1.63 (d)) *(for continuation/divisional with Box 17 completed)*
      i. [ ] DELETION OF INVENTOR(S) Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1 63(d)(2) and 1 33(b)
6. [ ] Application Data Sheet. See 37 CFR 1.76
7. [ ] CD-ROM or CD-R in duplicate, large table or Computer Program *(Appendix)*
8. Nucleotide and/or Amino Acid Sequence Submission *(if applicable, all necessary)*
   a. [ ] Computer Readable Form (CRF)
   b. Specification Sequence Listing on:
      i. [ ] CD-ROM or CD-R (2 copies); or
      i i. [ ] paper
   c. [ ] Statements verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

9. [ ] Assignment Papers (cover sheet & document(s))
10. [ ] 37 CFR 3.73(b) Statement *(when there is an assignee)* [ ] Power of Attorney
11. [ ] English Translation Document *(if applicable)*
12. [ ] Information Disclosure Statement (IDS)/PTO-1449 [ ] Copies of IDS Citations
13. [ ] Preliminary Amendment
14. [XX] Return Receipt Postcard (MPEP 503) *(Should be specifically itemized)*
15. [ ] Certified Copy of Priority Document(s) *(if foreign priority is claimed)*
16. [XX] Other: List of Inventors

17. If a CONTINUING APPLICATION, *check appropriate box, and supply the requisite information below and in a preliminary amendment, or in an Application Data Sheet under 37 CFR 1.76:*
    [ ] Continuation [ ] Divisional [ ] Continuation-in-part (CIP) of prior application No.: _____/_____
    Prior application information: *Examiner* _____ *Group / Art Unit:* _____

For CONTINUATION OR DIVISIONAL APPS only: The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 5b, is considered a part of the disclosure of the accompanying continuation or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.

18. CORRESPONDENCE ADDRESS

[XX] *Customer Number or Bar Code Label* 25748 *(Insert Customer No. or Attach bar code label here)* or [ ] *Correspondence address below*

Name					
Address					
City		State		Zip Code	
Country		Telephone		Fax	

Name (Print/Type)	Michael J. Schmelzer	Registration No. (Attorney/Agent)	43,093
Signature		Date	Nov. 9, 2000

Burden Hour Statement: This form is estimated to take 0 2 hours to complete Time will vary depending upon the needs of the individual case Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231 DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231.

List of Inventors

Steven Istvan LADUNGA
209 Sandpiper Court
Foster City, CA 94404-1320
Citizenship: HU

Maureen HIGGINS
5920 Conway Rd.
Bethesda, MD 20817
Citizenship: US

ISOLATED HUMAN SECRETED PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND USES THEREOF

RELATED APPLICATIONS

The present application claims priority to provisional application U.S. Serial No. (To Be Assigned), filed October 27, 2000 (Atty. Docket CL000841-PROV).

FIELD OF THE INVENTION

The present invention is in the field of secreted proteins that are related to the retinoic acid receptor responder secreted subfamily, recombinant DNA molecules, and protein production. The present invention specifically provides novel peptides and proteins that effect protein phosphorylation and nucleic acid molecules encoding such peptide and protein molecules, all of which are useful in the development of human therapeutics and diagnostic compositions and methods.

BACKGROUND OF THE INVENTION

Secreted Proteins

Many human proteins serve as pharmaceutically active compounds. Several classes of human proteins that serve as such active compounds include hormones, cytokines, cell growth factors, and cell differentiation factors. Most proteins that can be used as a pharmaceutically active compound fall within the family of secreted proteins. It is, therefore, important in developing new pharmaceutical compounds to identify secreted proteins that can be tested for activity in a variety of animal models. The present invention advances the state of the art by providing many novel human secreted proteins.

Secreted proteins are generally produced within cells at rough endoplasmic reticulum, are then exported to the golgi complex, and then move to secretory vesicles or granules, where they are secreted to the exterior of the cell via exocytosis.

Secreted proteins are particularly useful as diagnostic markers. Many secreted proteins are found, and can easily be measured, in serum. For example, a 'signal sequence trap' technique can often be utilized because many secreted proteins, such as certain secretory breast cancer proteins, contain a molecular signal sequence for cellular export. Additionally, antibodies against particular secreted serum proteins can serve as potential diagnostic agents, such as for diagnosing cancer.

Secreted proteins play a critical role in a wide array of important biological processes in humans and have numerous utilities; several illustrative examples are discussed herein. For example, fibroblast secreted proteins participate in extracellular matrix formation. Extracellular matrix affects growth factor action, cell adhesion, and cell growth. Structural and quantitative characteristics of fibroblast secreted proteins are modified during the course of cellular aging and such aging related modifications may lead to increased inhibition of cell adhesion, inhibited cell stimulation by growth factors, and inhibited cell proliferative ability (Eleftheriou *et al.*, *Mutat Res* 1991 Mar-Nov;256(2-6):127-38).

The secreted form of amyloid beta/A4 protein precursor (APP) functions as a growth and/or differentiation factor. The secreted form of APP can stimulate neurite extension of cultured neuroblastoma cells, presumably through binding to a cell surface receptor and thereby triggering intracellular transduction mechanisms. (Roch *et al.*, *Ann N Y Acad Sci* 1993 Sep 24;695:149-57). Secreted APPs modulate neuronal excitability, counteract effects of glutamate on growth cone behaviors, and increase synaptic complexity. The prominent effects of secreted APPs on synaptogenesis and neuronal survival suggest that secreted APPs play a major role in the process of natural cell death and, furthermore, may play a role in the development of a wide variety of neurological disorders, such as stroke, epilepsy, and Alzheimer's disease (Mattson *et al.*, *Perspect Dev Neurobiol* 1998; 5(4):337-52).

Breast cancer cells secrete a 52K estrogen-regulated protein (see Rochefort *et al.*, *Ann N Y Acad Sci* 1986;464:190-201). This secreted protein is therefore useful in breast cancer diagnosis.

Two secreted proteins released by platelets, platelet factor 4 (PF4) and beta-thromboglobulin (betaTG), are accurate indicators of platelet involvement in hemostasis

and thrombosis and assays that measure these secreted proteins are useful for studying the pathogenesis and course of thromboembolic disorders (Kaplan, *Adv Exp Med Biol* 1978;102:105-19).

Vascular endothelial growth factor (VEGF) is another example of a naturally secreted protein. VEGF binds to cell-surface heparan sulfates, is generated by hypoxic endothelial cells, reduces apoptosis, and binds to high-affinity receptors that are up-regulated by hypoxia (Asahara *et al.*, *Semin Interv Cardiol* 1996 Sep;1(3):225-32).

Many critical components of the immune system are secreted proteins, such as antibodies, and many important functions of the immune system are dependent upon the action of secreted proteins. For example, Saxon *et al.*, *Biochem Soc Trans* 1997 May;25(2):383-7, discusses secreted IgE proteins.

For a further review of secreted proteins, see Nilsen-Hamilton *et al., Cell Biol Int Rep* 1982 Sep;6(9):815-36.

Retinoic Acids and Retinoic Acid Receptors

Retinoids, or vitamin A metabolites/derivatives, have been determined to play essential roles in many aspects of development, metabolism and reproduction in vertebrates (see, for example, The Retinoids, Second Edition, Sporn et al. (Raven Press, New York, 1994)). There are two classes of retinoid receptors: the retinoic acid receptors (RARs), which bind to both all-trans retinoic acid (atRA) and 9-cis retinoic acid (9cRA), and the retinoid X receptors (RXRs), which bind only to 9cRA. These receptors modulate ligand-dependent gene expression by interacting as RXR/RAR heterodimers or RXR homodimers on specific target gene DNA sequences known as hormone response elements. In addition to their role in retinoid signalling, RXRs also serve as heterodimeric partners of nuclear receptors for vitamin D, thyroid hormone, and peroxisome proliferators (reviewed by Mangelsdorf et al., at pages 319-349 of The Retinoids, Second Edition, Sporn et al. (Raven Press, New York, 1994)).

A number of studies have demonstrated that retinoids are essential for normal growth, vision, tissue homeostasis, reproduction and overall survival (for reviews and references, See Sporn et al., The Retinoids, Vols. 1 and 2, Sporn et al., eds., Academic Press, Orlando, Fla. (1984)). For example, retinoids have been shown to be vital to the maintenance of skin homeostasis and barrier function in mammals (Fisher, G. J., and

Voorhees, J. J., FASEB J. 10:1002-1013 (1996)). Retinoids are also apparently crucial during embryogenesis, since offspring of dams with vitamin A deficiency (VAD) exhibit a number of developmental defects (Wilson, J. G., et al., Am. J. Anat. 92:189-217 (1953); Morriss-Kay, G. M., and Sokolova, N., FASEB J. 10:961-968 (1996)). With the exceptions of those on vision (Wald, G., et al., Science 162:230-239 (1968)) and spermatogenesis in mammals (van Pelt, H. M. M., and De Rooij, D. G., Endocrinology 128:697-704 (1991)), most of the effects generated by VAD in animals and their fetuses can be prevented and/or reversed by retinoic acid (RA) administration (Wilson, J. G., et al., Am. J. Anat. 92:189-217 (1953); Thompson et al., Proc. Royal Soc. 159:510-535 (1964); Morriss-Kay, G. M., and Sokolova, N., FASEB J. 10:961-968 (1996)). The dramatic teratogenic effects of maternal RA administration on mammalian embryos (Shenefelt, R. E., Teratology 5, 103-108 (1972); Kessel, M., Development 115:487-501 (1992); Creech Kraft, J., In Retinoids in Normal Development and Teratogenesis, G. M. Morriss-Kay, ed., Oxford University Press, Oxford, UK, pp. 267-280 (1992)), and the marked effects of topical administration of retinoids on embryonic development of vertebrates and limb regeneration in amphibians (Mohanty-Hejmadi et al., Nature 355:352-353 (1992); Tabin, C. J., Cell 66:199-217 (1991)), have contributed to the notion that RA may have critical roles in morphogenesis and organogenesis.

Many synthetic structural analogues of all-trans retinoic acid or 9-cis-retinoic acid, commonly termed "retinoids", have been described in the literature to date. Some of these molecules are able to bind to, and specifically activate, the RARs or, on the other hand, the RXRs. Furthermore, some analogues are able to bind to, and activate a particular RAR receptor subtype (.alpha., .beta. or .gamma.). Finally, other analogues do not exhibit any particular selective activity with regard to these different receptors. In this respect, and by way of example, 9-cis-retinoic acid activates the RARs and the RXRs at one and the same time without any noteworthy selectivity for either of these receptors (nonspecific agonist ligand), whereas all-trans retinoic acid selectively activates the RARs (RAR-specific agonist ligand), with all subtypes being included. In a general manner, and qualitatively, a given substance (or ligand) is said to be specific for a given family of receptors (or, respectively, for a particular receptor of this family) when the said substance exhibits an affinity for all the receptors of this family (or, respectively, for

the particular receptor of this family) which is stronger than that which it otherwise exhibits for all the receptors of any other family (or, respectively, for all the other receptors, of this same family or not).

The genetic activities of the RA signal are mediated through the two families of receptors--the RAR family and the RXR family--which belong to the superfamily of ligand-inducible transcriptional regulatory factors that include steroid/thyroid hormone and vitamin D3 receptors (for reviews see Leid et al., TIBS 17:427-433 (1992); Chambon, P., Semin. Cell Biol. 5:115-125 (1994); Chambon, P., FASEB J. 10:940-954 (1996); Giguere, V., Endocrinol. Rev. 15:61-79 (1994); Mangelsdorf, D. J., and Evans, R. M., Cell 83:841-850 (1995); Gronemeyer, H., and Laudet, V., Protein Profile 2:1173-1236 (1995)).

RARs are the critical factors in tissue differentiation and development. They are up-regulated in rapidly dividing cells and tumors. RARs play an important role in lymphocyte activation. Synthetic antagonists of retinoic acid receptors can inhibit delayed type hypersensitivity (DTH). Growth factors and carotene regulate RXR expression levels. For example, granulocyte macrophage colony-stimulating factor induces retinoic acid receptors in myeloid leukemia cells.

Retinoic acid receptors can form heterodimers with other nuclear receptors. The protein provided by the present invention can be used as a probe to detect possible interactions in the two-hybrid assay. Synthetic peptides that mimic dimerization surface can disrupt intermolecular interactions between these receptors. RAR gene rearrangements are the primary causes of some types of leukemia and provide a convenient genetic marker for malignant cell lines. A number of retinoic acid derivatives are used in treatment of myelodysplastic disorders. They are designed to bind and activate RXRs. Beta-carotene can prevent skin tumor formation in mouse models. N-(4-hydroxyphenyl)retinamide can delay onset of dysplasia in bronchi. Different chemopreventive drugs can be designed to target individual retinoic receptors. The sequences provided by the present invention may be used to design high affinity chemopreventive compounds.

Although both the RARs and RXRs respond to all-trans-retinoic acid in vivo, the receptors differ in several important aspects. First, the RARs and RXRs are significantly

divergent in primary structure (e.g., the ligand binding domains of RAR.alpha. and RXR.alpha. have only 27% amino acid identity). These structural differences are reflected in the different relative degrees of responsiveness of RARs and RXRs to various vitamin A metabolites and synthetic retinoids. In addition, distinctly different patterns of tissue distribution are seen for RARs and RXRs. For example, in contrast to the RARs, which are generally not expressed at high levels in the visceral tissues, RXR.alpha. mRNA has been shown to be most abundant in the liver, kidney, lung, muscle and intestine. Finally, the RARs and RXRs have different target gene specificity. For example, response elements have recently been identified in the cellular retinal binding protein type II (CRBPII) and apolipoprotein AI genes which confer responsiveness to RXR, but not RAR. Furthermore, RAR has also been recently shown to repress RXR-mediated activation through the CRBPII RXR response element (Manglesdorf et al., Cell, 66:555-61 (1991)). These data indicate that two retinoic acid responsive pathways are not simply redundant, but instead manifest a complex interplay. Recently, Heyman et al. (Cell, 68:397-406 (1992)) and Levin et al. (Nature, 355:359-61 (1992)) independently demonstrated that 9-cis-retinoic acid is a natural endogenous ligand for the RXRs. 9-cis-retinoic acid was shown to bind and transactivate the RXRs, as well as the RARs, and therefore appears to act as a "bifunctional" ligand.

RAR Receptors

Receptors belonging to the RAR family (RAR.alpha., .beta. and .gamma. and their isoforms) are activated by both all-trans- and 9-cis-RA (Leid et al., TIBS 17:427-433 (1992); Chambon, P., Semin. Cell Biol. 5:115-125 (1994); Dolle, P., et al., Mech. Dev. 45:91-104 (1994); Chambon, P., FASEB J. 10:940-954 (1996)). Within a given species, the DNA binding (C) and the ligand binding (E) domains of the three RAR types are highly similar, whereas the C-terminal domain F and the middle domain D exhibit no or little similarity. The amino acid sequences of the three RAR types are also notably different in their B regions, and their main isoforms (.alpha.1 and .alpha.2, .beta.1 to .beta.4, and .gamma.1 and .gamma.2) further differ in their N-terminal A regions (Leid et al., TIBS 17:427-433 (1992)). Amino acid sequence comparisons have revealed that the interspecies conservation of a given RAR type is greater than the similarity found between the three RAR types within a given species (Leid et al., TIBS 17:427-433

(1992)). This interspecies conservation is particularly striking in the N-terminal A regions of the various RAR.alpha., .beta. and .gamma. isoforms, whose A region amino acid sequences are quite divergent. Taken together with the distinct spatio-temporal expression patterns observed for the transcripts of each RAR and RXR type in the developing embryo and in various adult mouse tissues (Zelent, A., et al., Nature 339:714-717 (1989); Dolle, P., et al., Nature 342:702-705 (1989); Dolleet al., Development 110:1133-1151 (1990); Ruberte et al., Development 108:213-222 (1990); Ruberte et al., Development 111:45-60 (1991); Mangelsdorf et al., Genes & Dev. 6:329-344 (1992)), this interspecies conservation has suggested that each RAR type (and isoform) may perform unique functions. This hypothesis is further supported by the finding that the various RAR isoforms contain two transcriptional activation functions (AFs) located in the N-terminal A/B region (AF-1) and in the C-terminal E region (AF-2), which can synergistically, and to some extent differentially, activate various RA-responsive promoters (Leid et al., TIBS 17:427-433 (1992); Nagpal, S., et al., Cell 70:1007-1019 (1992); Nagpal, S., et al., EMBO J. 12:2349-2360 (1993)).

RXR Receptors

Unlike the RARs, members of the retinoid X receptor family (RXR.alpha., .beta. and .gamma.) are activated exclusively by 9-cis-RA (Chambon, P., FASEB J. 10:940-954 (1996); Chambon, P., Semin. Cell Biol. 5:115-125 (1994); Dolle, P., et al., Mech. Dev. 45:91-104 (1994); Linney, E., Current Topics in Dev. Biol. 27:309-350 (1992); Leid et al., TIBS 17:427-433 (1992); Kastner et al., in Vitamin A in Health and Disease, R. Blomhoff, ed., Marcel Dekker, New York (1993)). However, the RXRs characterized to date are similar to the RARs in that the different RXR types also differ markedly in their N-terminal A/B regions (Leid et al., TIBS 17:427-433 (1992); Leid et al., Cell 68:377-395 (1992); Mangelsdorf et al., Genes and Dev. 6:329-344 (1992)), and contain the same transcriptional activation functions in their N-terminal A/B region and C-terminal E region (Leid et al., TIBS 17:427-433 (1992); Nagpal, S., et al., Cell 70:1007-1019 (1992); Nagpal, S., et al., EMBO J. 12:2349-2360 (1993)).

RXR.alpha. and RXR.beta. have a widespread (possibly ubiquitous) expression pattern during mouse development and in the adult animal, being found in all fetal and adult tissues thus far examined (Mangelsdorf, D. J., et al., Genes & Devel. 6:329-344

(1992); Dolle, P., et al., Mech. Devel. 45:91-104 (1994); Nagata, T., et al., Gene 142:183-189 (1994)). RXR.gamma. transcripts, however, appear to have a more restricted distribution, being expressed in developing skeletal muscle in the embryo (where their expression persists throughout life), in the heart (after birth), in sensory epithelia of the visual and auditory systems, in specific structures of the central nervous system, and in tissues involved in thyroid hormone homeostasis, e.g., the thyroid gland and thyrotrope cells in the pituitary (Mangelsdorf, D. J., et al., Genes & Devel. 6:329-344 (1992); Dolle, P., et al., Mech. Devel. 45:91-104 (1994); Sugawara, A., et al., Endocrinology 136:1766-1774 (1995); Liu, Q., and Linney, E., Mol. Endocrinol. 7:651-658 (1993)).

It is currently unclear whether all the molecular properties of RXRs characterized in vitro are relevant for their physiological functions in vivo. In particular, it is unknown under what conditions these receptors act as 9-cis-RA-dependent transcriptional regulators (Chambon, P., Semin. Cell Biol. 5:115-125 (1994)). The knock-outs of RXR.alpha. and RXR.beta. in the mouse have provided some insight into the physiological functions of these receptors. For example, the ocular and cardiac malformations observed in RXR.alpha..sup.-/- fetuses (Kastner, P., et al., Cell 78:987-1003 (1994); Sucov, H. M., et al., Genes & Devel. 8:1007-1018 (1994)) are similar to those found in the fetal VAD syndrome, thus suggesting an important function of RXR.alpha. in the transduction of a retinoid signal during development. The involvement of RXRs in retinoid signaling is further supported by studies of compound RXR.alpha./RAR mutants, which reveal defects that are either absent or less severe in the single mutants (Kastner, P., et al., Cell 78:987-1003 (1994); Kastner, P., et al., Cell 83:859-869 (1995)). Interestingly, however, knockout of RXR.gamma. in the mouse induces no overt deleterious effects, and RXR.gamma..sup.-/- homozygotes which are also RXR.alpha..sup.-/- or RXR.beta..sup.-/- exhibit no additional abnormalities beyond those seen in RXR.alpha..sup.-/-, RXR.beta..sup.-/- and fetal VAD syndrome fetuses (Krezel, W., et al., Proc. Natl. Acad. Sci. USA 93(17):9010-9014 (1996)), suggesting that RXR.gamma., despite its highly tissue-specific expression pattern in the developing embryo, is dispensable for embryonic development and postnatal life in the mouse. The observation that live-born RXR.gamma..sup.-/- /RXR.beta..sup.-/- /RXR.alpha..sup.-/- mutants can grow to reach adult age (Krezel et al., Proc. Natl. Acad. Sci. USA

93(17):9010-9014 (1996)) indicates that a single RXR.alpha. allele is sufficient to carry out all of the vital developmental and postnatal functions of the RXR family of receptors, particularly all of the developmental functions which depend on RARs and may require RXR partnership (Dolle, P., et al., Mech. Dev. 45:91-104 (1994); Kastner, P., et al., Cell 83:859-869 (1995)). Furthermore, the finding that RXR.alpha..sup.-/- /RXR.gamma..sup.-/- double mutant embryos are not more affected than are single RXR.alpha..sup.-/- mutants (Krezel et al., Proc. Natl. Acad. Sci. USA 93(17):9010-9014 (1996)) clearly shows that RXR.beta. alone can also perform some of these functions. Therefore, the fact that RXR.alpha. alone and, to a certain extent RXR.beta. alone, are sufficient for the completion of a number of developmental RXR functions, clearly indicates the existence of a large degree of functional redundancy amongst RXRs. In this respect, the RXR situation is different from that of RARs, since all of types of RAR double mutants displayed much broader sets of defects than single mutants (Rowe, A., et al., Develop. 111:771-778 (1991); Lohnes, D., et al., Develop. 120:2723-2748 (1994); Mendelsohn, C., Develop. 120:2749-2771 (1994)).

Retinoid Binding to RAR and RXR Receptors

The crystal structures of the ligand-binding domains (LBDs) of the RARs and RXRs have recently been elucidated (Bourget, W., et al., Nature 375:377-382 (1995); Renaud, J. P., et al., Nature 378:681-689 (1995); Wurtz, J. M., et al., Nature Struct. Biol. 3:87-94 (1996)). Among the various RAR types, substantial amino acid sequence identity is observed in these domains: comparison of the LBDs of RAR.alpha., RAR.beta. and RAR.gamma. indicates that only three amino acid residues are variable in the ligand-binding pocket of these receptors. These residues apparently account for the fact that the various RAR types exhibit some selectivity in binding certain synthetic retinoids (Chen, J. -Y., et al., EMBO J. 14(6):1187-1197 (1995); Renaud, J. P., et al., Nature 378:681-689 (1995)), and consideration of these divergent residues can be used to design RAR type-specific synthetic retinoids which may be agonistic or antagonistic (Chambon, P., FASEB J. 10:940-954 (1996)). This design approach may be extendable generally to other nuclear receptors, such as thyroid receptor .alpha. (Wagner, R. L., et al., Nature 378:690-697 (1995)), the ligand-binding pockets of which may chemically and structurally resemble those of the RARs (Chambon, P., FASEB J. 10:940-954 (1996)). Conversely,

molecular modeling of the ligand-binding pocket of the RXRs demonstrates that there are no overt differences in amino acid composition between RXR.alpha., RXR.beta. and RXR.gamma. (Bourguet, W., et al., Nature 375:377-382 (1995); Wurtz, J. M., et al., Nature Struct. Biol. 3:87-94 (1996)), suggesting that design of type-specific synthetic ligands for the RXRs may be more difficult than for the RARs (Chambon, P., FASEB J. 10:940-954 (1996)).

Retinoid Signaling Through RAR:RXR Heterodimers

Nuclear receptors (NRs) are members of a superfamily of ligand-inducible transcriptional regulatory factors that include receptors for steroid hormones, thyroid hormones, vitamin D3 and retinoids (Leid, M., et al., Trends Biochem. Sci. 17:427-433 (1992); Leid, M., et al., Cell 68:377-395 (1992); and Linney, E. Curr. Top. Dev. Biol., 27:309-350 (1992)). NRs exhibit a modular structure which reflects the existence of several autonomous functional domains. Based on amino acid sequence similarity between the chicken estrogen receptor, the human estrogen and glucocorticoid receptors, and the v-erb-A oncogene (Krust, A., et al., EMBO J. 5:891-897 (1986)), defined six regions--A, B, C, D, E and F--which display different degrees of evolutionary conservation amongst various members of the nuclear receptor superfamily. The highly conserved region C contains two zinc fingers and corresponds to the core of the DNA-binding domain (DBD), which is responsible for specific recognition of the cognate response elements. Region E is functionally complex, since in addition to the ligand-binding domain (LBD), it contains a ligand-dependent activation function (AF-2) and a dimerization interface. An autonomous transcriptional activation function (AF-1) is present in the non-conserved N-terminal A/B regions of the steroid receptors. Interestingly, both AF-1 and AF-2 of steroid receptors exhibit differential transcriptional activation properties which appear to be both cell type and promoter context specific (Gronemeyer, H. Annu. Rev. Genet. 25:89-123 (1991)).

As described above, the all-trans (T-RA) and 9-cis (9C-RA) retinoic acid signals are transduced by two families of nuclear receptors, RAR .alpha., .beta. and .gamma. (and their isoforms) are activated by both T-RA and 9C-RA, whereas RXR .alpha., .beta. and .gamma. are exclusively activated by 9C-RA (Allenby, G. et al., Proc. Natl. Acad. Sci. USA 90:30-34 (1993)). The three RAR types differ in their B regions, and their main

isoforms (.alpha.1 and .alpha.2, .beta.1-4, and .gamma.1 and .gamma.2) have different N-terminal A regions (Leid, M. et al., Trends Biochem. Sci. 17:427-433 (1992)). Similarly, the RXR types differ in their A/B regions (Mangelsdorf, D. J. et al., Genes Dev. 6:329-344 (1992)).

The E-region of RARs and RXRs has also been shown to contain a dimerization interface (Yu, V. C. et al., Curr. Opin. Biotechnol. 3:597-602 (1992)). Most interestingly, it was demonstrated that RAR/RXR heterodimers bind much more efficiently in vitro than homodimers of either receptor to a number of RA response elements (RAREs, also known as retinoic acid receptor responders) (Yu, V. C. et al., Cell 67:1251-1266 (1991); Berrodin, T. J. et al., Mol. Endocrinol 6:1468-1478 (1992); Bugge, T. H. et al., EMBO J. 11:1409-1418 (1992); Hall, R. K. et al., Mol. Cell. Biol. 12: 5527-5535 (1992); Hallenbeck P. L. et al., Proc. Natl. Acad. Sci. USA 89:5572-5576 (1992); Husmann, M. et al., Biochem. Biophys. Res. Commun. 187:1558-1564 (1992); Kliewer, S. A. et al., Nature 355:446-449 (1992); Leid, M. et al., Cell 68:377-395 (1992); Marks, M. S. et al., EMBO J. 11:1419-1435 (1992); Zhang, X. K. et al., Nature 355:441-446 (1992)). RAR and RXR heterodimers are also preferentially formed in solution in vitro (Yu, V. C. et al., Cell 67:1251-1266 (1991); Leid, M. et al., Cell 68:377-395 (1992); Marks, M. S. et al., EMBO J. 11:1419-1435 (1992)), although the addition of 9C-RA appears to enhance the formation of RXR homodimers in vitro (Lehman, J. M. et al., Science 258:1944-1946 (1992); Zhang, X. K. et al., Nature 358:587-591 (1992b)).

It has been shown that activation of RA-responsive promoters likely occurs through RAR:RXR heterodimers rather than through homodimers (Yu, V. C. et al., Cell 67:1251-1266 (1991); Leid et al., Cell 68:377-395 (1992b); Durand et al., Cell 71:73-85 (1992); Nagpal et al., Cell 70:1007-1019 (1992); Zhang, X. K., et al., Nature 355, 441-446 (1992); Kliewer et al., Nature 355:446-449 (1992); Bugge et al., EMBO J. 11: 1409-1418 (1992); Marks et al., EMBO J. 11:1419-1435 (1992); Yu, V. C. et al., Cur. Op. Biotech. 3:597-602 (1992); Leid et al., TIBS 17:427-433 (1992); Laudet and Stehelin, Curr. Biol. 2:293-295 (1992); Green, S., Nature 361:590-591 (1993)). The RXR portion of these heterodimers has been proposed to be silent in retinoid-induced signaling (Kurokawa, R., et al., Nature 371:528-531 (1994); Forman, B. M., et al., Cell 81:541-550 (1995); Mangelsdorf, D. J., and Evans, R. M., Cell 83:835-850 (1995)), although

conflicting results have been reported on this issue (Apfel, C. M., et al., J. Biol. Chem. 270(51):30765-30772 (1995); see Chambon, P., FASEB J. 10:940-954 (1996) for review). Although the results of these studies strongly suggest that RAR/RXR heterodimers are indeed functional units that transduce the RA signal in vivo, it is unclear whether all of the suggested heterodimeric combinations occur in vivo (Chambon, P., Semin. Cell Biol. 5:115-125 (1994)). Thus, the basis for the highly pleiotropic effect of retinoids may reside, at least in part, in the control of different subsets of retinoid-responsive promoters by cell-specifically expressed heterodimeric combinations of RAR:RXR types (and isoforms), whose activity may be in turn regulated by cell-specific levels of all-trans- and 9-cis-RA (Leid et al., TIBS 17:427-433 (1992)).

The RXR receptors may also be involved in RA-independent signaling. For example, the observation of aberrant lipid metabolism in the Sertoli cells of RXR.beta..sup.-/- mutant animals suggests that functional interactions may also occur between RXR.beta. and the peroxisomal proliferator-activated receptor signaling pathway (WO 94/26100; Kastner, P., et al., Genes & Devel. 10:80-92 (1996)).

For a further review of retinoic acid receptors, see: Shimizu et al., Cancer Res 2000 Aug 15;60(16):4544-9; Ponnamperuma et al., Nutr Cancer 2000;37(1):82-8; Yoshimura et al., J Med Chem 2000 Jul 27;43(15):2929-37; Kurie et al., Clin Cancer Res 2000 Aug;6(8):2973-9; Lee et al., J Biol Chem 2000 Aug 17; and Sainty et al., Blood 2000 Aug 15;96(4):1287-96.

Secreted proteins, particularly members of the retinoic acid receptor responder secreted protein subfamily, are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown members of this subfamily of secreted proteins. The present invention advances the state of the art by providing previously unidentified human secreted proteins that have homology to members of the retinoic acid receptor responder secreted protein subfamily.

SUMMARY OF THE INVENTION

The present invention is based in part on the identification of amino acid sequences of human secreted peptides and proteins that are related to the retinoic acid

receptor responder secreted protein subfamily, as well as allelic variants and other mammalian orthologs thereof. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents that modulate secreted protein activity in cells and tissues that express the secreted protein.

DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequence of a cDNA molecule or transcript sequence that encodes the secreted protein of the present invention. (SEQ ID NO:1) In addition, structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of inventions based on this molecular sequence.

FIGURE 2 provides the predicted amino acid sequence of the secreted protein of the present invention. (SEQ ID NO:2) In addition structure and functional information such as protein family, function, and modification sites is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence.

FIGURE 3 provides genomic sequences that span the gene encoding the secreted protein of the present invention. (SEQ ID NO:3) In addition structure and functional information, such as intron/exon structure, promoter location, etc., is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence. As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 16.

DETAILED DESCRIPTION OF THE INVENTION

General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information

revealed previously unidentified fragments of the human genome that encode peptides that share structural and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a secreted protein or part of a secreted protein and are related to the retinoic acid receptor responder secreted protein subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of human secreted peptides and proteins that are related to the retinoic acid receptor responder secreted protein subfamily, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these secreted peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the secreted protein of the present invention.

In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known secreted proteins of the retinoic acid receptor responder secreted protein subfamily and the expression pattern observed. The art has clearly established the commercial importance of members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the known retinoic acid receptor responder family or subfamily of secreted proteins.

Specific Embodiments

Peptide Molecules

The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the secreted protein family of proteins and are related to the retinoic acid receptor responder secreted protein subfamily

(protein sequences are provided in Figure 2, transcript/cDNA sequences are provided in Figure 1 and genomic sequences are provided in Figure 3). The peptide sequences provided in Figure 2, as well as the obvious variants described herein, particularly allelic variants as identified herein and using the information in Figure 3, will be referred herein as the secreted peptides of the present invention, secreted peptides, or peptides/proteins of the present invention.

The present invention provides isolated peptide and protein molecules that consist of, consist essentially of, or comprise the amino acid sequences of the secreted peptides disclosed in the Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA or Figure 3, genomic sequence), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the secreted peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less

than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

The isolated secreted peptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. For example, a nucleic acid molecule encoding the secreted peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). The amino acid sequence of such a protein is provided in Figure 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that comprise the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or

more additional amino acids. The preferred classes of proteins that are comprised of the secreted peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The secreted peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a secreted peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the secreted peptide. "Operatively linked" indicates that the secreted peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the secreted peptide.

In some uses, the fusion protein does not affect the activity of the secreted peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant secreted peptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A secreted peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the secreted peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as

naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the secreted peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York,

1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al., Nucleic Acids Res. 12(1)*:387 (1984)) (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the secreted peptides of the present invention as well as being encoded by the same genetic locus as the secreted peptide provided herein.

Allelic variants of a secreted peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the secreted peptide as well as being encoded by the same genetic locus as the secreted peptide provided herein. Genetic locus can readily be determined based on the genomic information provided in Figure 3, such as the genomic sequence mapped to the reference human. As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 16. As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a secreted peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Figure 3 provides SNP information that has been found in a gene encoding the transporter proteins of the present invention. The following variations were identified: T948-, G1149A, G4199C, T4352G, A6493G, T14047GA, G14136T, G14238A, T14260G, C25736T, G26321A, A31359G, T35098G, C40532G, A41706G, G51095C, G53101C, G54556C, - 61872T, G62172C, A62860C, C67086A, A67621C, A70582T, C74175-, T74478C, T77092G, T77328C, G77385A, C77947T, G79395C, -81111C, and G81610A.

Paralogs of a secreted peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the secreted peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a secreted peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a secreted peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the secreted peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a secreted peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the secreted peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the secreted peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a secreted peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science 247*:1306-1310 (1990).

Variant secreted peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind substrate, ability to phosphorylate substrate, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 2 provides the result of protein analysis and can be used to identify critical domains/regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science 244*:1081-1085 (1989)), particularly using the results provided in Figure 2. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as secreted protein activity or in assays such as an *in vitro* proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al., J. Mol. Biol. 224*:899-904 (1992); de Vos *et al. Science 255*:306-312 (1992)).

The present invention further provides fragments of the secreted peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly those comprising the residues identified in Figure 2. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16, or more contiguous amino acid residues from a secreted peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the secreted peptide or could be chosen for the ability to perform a function, e.g. bind a substrate or act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the secreted peptide, e.g., active site or a substrate-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in Figure 2.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in secreted peptides are described in

basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art (some of these features are identified in Figure 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol. 182*: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci. 663*:48-62 (1992)).

Accordingly, the secreted peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature secreted peptide is fused with another compound, such as a compound to increase the half-life of the secreted peptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature secreted peptide, such as a leader or secretory sequence or a sequence for purification of the mature secreted peptide or a pro-protein sequence.

Protein/Peptide Uses

The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or ligand) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein or ligand (such as, for example, in a secreted protein-effector protein interaction or secreted protein-ligand interaction), the protein can be used to identify the binding partner/ligand so as to develop a system to identify inhibitors of the binding interaction. Any or all of these uses are capable of being developed into reagent grade or kit format for commercialization as commercial products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, secreted proteins isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug, particularly in modulating a biological or pathological response in a cell or tissue that expresses the secreted protein. A large percentage of pharmaceutical agents are being developed that modulate the activity of secreted proteins, particularly members of the retinoic acid receptor responder subfamily (see Background of the Invention). The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention, particularly in combination with the expression information provided in Figure 1. Such uses can readily be determined using the information provided herein, that which is known in the art, and routine experimentation.

The proteins of the present invention (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to secreted proteins that are related to members of the retinoic acid receptor responder subfamily. Such assays involve any of the known secreted protein functions or activities or properties useful for diagnosis and treatment of secreted protein-related conditions that are specific for the subfamily of secreted proteins that the one of the present invention belongs to, particularly in cells and tissues that express the secreted protein.

The proteins of the present invention are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the secreted protein, as a biopsy or expanded in cell culture. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the secreted protein.

The polypeptides can be used to identify compounds that modulate secreted protein activity of the protein in its natural state or an altered form that causes a specific disease or pathology associated with the secreted protein. Both the secreted proteins of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the secreted protein. These compounds can be further screened against a functional secreted protein to determine the effect of the compound on the secreted protein activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the secreted protein to a desired degree.

Further, the proteins of the present invention can be used to screen a compound for the ability to stimulate or inhibit interaction between the secreted protein and a molecule that normally interacts with the secreted protein, e.g. a substrate or a component of the signal pathway that the secreted protein normally interacts (for example, another secreted protein). Such assays typically include the steps of combining the secreted protein with a candidate compound under conditions that allow the secreted protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the secreted protein and the target.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature 354*:82-84 (1991); Houghten *et al.*, *Nature 354*:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell 72*:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, $F(ab')_2$, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the receptor that competes for substrate binding. Other candidate compounds include mutant secreted proteins or appropriate fragments containing mutations that affect secreted protein function and thus compete for substrate. Accordingly, a fragment that competes for substrate, for example with a higher affinity, or a fragment that binds substrate but does not allow release, is encompassed by the invention.

Any of the biological or biochemical functions mediated by the secreted protein can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the Figures, particularly Figure 2. Specifically, a biological function of a cell or tissues that expresses the secreted protein can be assayed.

Binding and/or activating compounds can also be screened by using chimeric secreted proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a substrate-binding region can be used that interacts with a different substrate then that which is recognized by the native secreted protein. Accordingly, a different set of signal transduction components is available as an end-point

assay for activation. This allows for assays to be performed in other than the specific host cell from which the secreted protein is derived.

The proteins of the present invention are also useful in competition binding assays in methods designed to discover compounds that interact with the secreted protein (e.g. binding partners and/or ligands). Thus, a compound is exposed to a secreted protein polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble secreted protein polypeptide is also added to the mixture. If the test compound interacts with the soluble secreted protein polypeptide, it decreases the amount of complex formed or activity from the secreted protein target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the secreted protein. Thus, the soluble polypeptide that competes with the target secreted protein region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the secreted protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ^{35}S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of secreted protein-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies

reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a secreted protein-binding protein and a candidate compound are incubated in the secreted protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the secreted protein target molecule, or which are reactive with secreted protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Agents that modulate one of the secreted proteins of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

Modulators of secreted protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the secreted protein pathway, by treating cells or tissues that express the secreted protein. These methods of treatment include the steps of administering a modulator of secreted protein activity in a pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described herein.

In yet another aspect of the invention, the secreted proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with the secreted protein and are involved in secreted protein activity.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a secreted protein is fused to a gene encoding the DNA binding domain of a known

transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming a secreted protein-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the secreted protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a secreted protein-modulating agent, an antisense secreted protein nucleic acid molecule, a secreted protein-specific antibody, or a secreted protein-binding partner) can be used in an animal or other model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The secreted proteins of the present invention are also useful to provide a target for diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism. The method involves contacting a biological sample with a compound capable of interacting with the secreted protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered secreted protein activity in cell-based or cell-free assay, alteration in substrate or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagent, such as an antibody or protein binding agent. Alternatively, the peptide can be detected *in vivo* in a subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual

permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the secreted protein in which one or more of the secreted protein functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other substrate-binding regions that are more or less active in substrate binding, and secreted protein activation. Accordingly, substrate dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein. Accordingly, methods for treatment include the use of the secreted protein or fragments.

Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in

response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or $F(ab')_2$, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, Antibodies, Cold Spring Harbor Press, (1989).

In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods and as presented in the Figures.

Antibodies are preferably prepared from regions or discrete fragments of the secreted proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or secreted protein/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2).

Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include

umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression,

antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the secreted peptide to a binding partner such as a substrate. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information relating to the proteins of the present invention.

The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nuleic acid arrays and similar methods have been developed for antibody arrays.

Nucleic Acid Molecules

The present invention further provides isolated nucleic acid molecules that encode a secreted peptide or protein of the present invention (cDNA, transcript and genomic

sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the secreted peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise the nucleotide sequences shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprises several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, humans genomic sequence (Figure 3) and cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the Figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking,

prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the secreted peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the secreted proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in Figures 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene

termination sequences. Such fragments are useful in controlling heterologous gene expression and in developing screens to identify gene-modulating agents. A promoter can readily be identified as being 5' to the ATG start site in the genomic sequence provided in Figure 3.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found

in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2.

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods.

The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention.

The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in secreted protein expression relative to normal results.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA include Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a secreted protein, such as by measuring a level of a secreted protein-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a secreted protein gene has been mutated.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate secreted protein nucleic acid expression.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the secreted protein gene, particularly biological and pathological processes that are mediated by the secreted protein in cells and tissues that express it. The method typically includes assaying the ability of the compound to modulate the expression of the secreted protein nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired secreted protein nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the secreted protein nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

Thus, modulators of secreted protein gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of secreted protein mRNA in the presence of the candidate compound is compared to the level of expression of secreted protein mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate secreted protein nucleic acid expression in cells and tissues that express the secreted protein. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or nucleic acid expression.

Alternatively, a modulator for secreted protein nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the secreted protein nucleic acid expression in the cells and tissues that express the protein.

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the secreted protein gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in secreted protein nucleic acid expression, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in secreted protein genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally occurring genetic mutations in the secreted protein gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the secreted protein gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a secreted protein.

Individuals carrying mutations in the secreted protein gene can be detected at the nucleic acid level by a variety of techniques. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science 241*:1077-1080 (1988); and Nakazawa *et al.*, *PNAS 91*:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the

gene (see Abravaya *et al.*, *Nucleic Acids Res. 23*:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a secreted protein gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant secreted protein gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C.W., (1995) *Biotechniques 19*:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr. 36*:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol. 38*:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science 230*:1242 (1985)); Cotton *et al.*, *PNAS 85*:4397 (1988); Saleeba *et al.*, *Meth. Enzymol. 217*:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS 86*:2766 (1989); Cotton *et al.*, *Mutat. Res. 285*:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl. 9*:73-79

(1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature 313*:495 (1985)). Examples of other techniques for detecting point mutations include selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the secreted protein gene in an individual in order to select an appropriate compound or dosage regimen for treatment.

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control secreted protein gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of secreted protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into secreted protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of secreted protein nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired secreted protein nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the secreted protein, such as substrate binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in secreted protein gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired secreted protein to treat the individual.

The invention also encompasses kits for detecting the presence of a secreted protein nucleic acid in a biological sample. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting secreted protein nucleic acid in a biological sample; means for determining the amount of secreted protein nucleic acid in the sample; and means for comparing the amount of secreted protein nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect secreted protein mRNA or DNA.

Nucleic Acid Arrays

The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in Figures 1 and 3 (SEQ ID NOS:1 and 3).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee *et al.*, PCT application W095/11995 (Chee *et al.*), Lockhart, D. J. *et al.* (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. *et al.* (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown *et al.*, US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in

length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a gene or genes of interest.

In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler *et al.*) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large-scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of the secreted proteins/peptides of the present invention. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes, at least one of which is a gene of the present invention and or alleles of the secreted protein gene of the present invention.

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the Human genome disclosed herein. Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques,* Elsevier

Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry,* Academic Press, Orlando, FL Vol. 1 (1 982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology,* Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the Human genome disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified secreted protein gene of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated

into one of the established kit formats which are well known in the art, particularly expression arrays.

Vectors/host cells

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, or MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in prokaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ, the lac, TRP, and TAC promoters

from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is

joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene 67*:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene 69*:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology 185*:60-89 (1990)).

Recombinant protein expression can be maximized in host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res. 20*:2111-2118 (1992)).

The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J. 6*:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell 30*:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene 54*:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol. 3*:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology 170*:31-39 (1989)).

In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature 329*:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J. 6*:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast,

other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell- free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as kinases, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the peptide is not secreted into the medium, which is typically the case with kinases, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a secreted protein or peptide that can be further purified to produce desired amounts of secreted protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

Host cells are also useful for conducting cell-based assays involving the secreted protein or secreted protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native secreted protein is useful for assaying compounds that stimulate or inhibit secreted protein function.

Host cells are also useful for identifying secreted protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the

mutant secreted protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native secreted protein.

Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a secreted protein and identifying and evaluating modulators of secreted protein activity. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the secreted protein nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the secreted protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues

in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al. PNAS 89*:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science 251*:1351-1355 (1991). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature 385*:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_o phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect substrate binding, secreted protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* secreted protein function, including substrate interaction, the effect of specific mutant secreted proteins on secreted protein function and substrate interaction, and the effect of chimeric secreted proteins. It is also possible to assess the

effect of null mutations, that is, mutations that substantially or completely eliminate one or more secreted protein functions.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

Claims

That which is claimed is:

1. An isolated peptide consisting of an amino acid sequence selected from the group consisting of:

(a) an amino acid sequence shown in SEQ ID NO:2;

(b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;

(c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3; and

(d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.

2. An isolated peptide comprising an amino acid sequence selected from the group consisting of:

(a) an amino acid sequence shown in SEQ ID NO:2;

(b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;

(c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3; and

(d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.

3. An isolated antibody that selectively binds to a peptide of claim 2.

4. An isolated nucleic acid molecule consisting of a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;

(b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;

(c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;

(d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and

(e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

5. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;

(b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;

(c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under

stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;

(d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and

(e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

6. A gene chip comprising a nucleic acid molecule of claim 5.

7. A transgenic non-human animal comprising a nucleic acid molecule of claim 5.

8. A nucleic acid vector comprising a nucleic acid molecule of claim 5.

9. A host cell containing the vector of claim 8.

10. A method for producing any of the peptides of claim 1 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.

11. A method for producing any of the peptides of claim 2 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.

12. A method for detecting the presence of any of the peptides of claim 2 in a sample, said method comprising contacting said sample with a detection agent that specifically allows detection of the presence of the peptide in the sample and then detecting the presence of the peptide.

13. A method for detecting the presence of a nucleic acid molecule of claim 5 in a sample, said method comprising contacting the sample with an oligonucleotide that hybridizes to said nucleic acid molecule under stringent conditions and determining whether the oligonucleotide binds to said nucleic acid molecule in the sample.

14. A method for identifying a modulator of a peptide of claim 2, said method comprising contacting said peptide with an agent and determining if said agent has modulated the function or activity of said peptide.

15. The method of claim 14, wherein said agent is administered to a host cell comprising an expression vector that expresses said peptide.

16. A method for identifying an agent that binds to any of the peptides of claim 2, said method comprising contacting the peptide with an agent and assaying the contacted mixture to determine whether a complex is formed with the agent bound to the peptide.

17. A pharmaceutical composition comprising an agent identified by the method of claim 16 and a pharmaceutically acceptable carrier therefor.

18. A method for treating a disease or condition mediated by a human secreted protein, said method comprising administering to a patient a pharmaceutically effective amount of an agent identified by the method of claim 16.

19. A method for identifying a modulator of the expression of a peptide of claim 2, said method comprising contacting a cell expressing said peptide with an agent, and determining if said agent has modulated the expression of said peptide.

20. An isolated human secreted peptide having an amino acid sequence that shares at least 70% homology with an amino acid sequence shown in SEQ ID NO:2.

21. A peptide according to claim 20 that shares at least 90 percent homology with an amino acid sequence shown in SEQ ID NO:2.

22. An isolated nucleic acid molecule encoding a human secreted peptide, said nucleic acid molecule sharing at least 80 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 or 3.

23. A nucleic acid molecule according to claim 22 that shares at least 90 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 or 3.

ABSTRACT

The present invention provides amino acid sequences of peptides that are encoded by genes within the human genome, the secreted peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogs of the secreted peptides, and methods of identifying modulators of the secreted peptides.

```
   1  ATGGGACCCT GGTCAGTGGT GGTGCTGGTG TGCGGCATGA AGCAGCTGGG
  51  GCAGGCCCTC CAGGCCTCAG TTTCCCTGTC TATCATCACA GAGAATCAGG
 101  GCAAAAGGTG TCCCTTCTGT GGAGCCCAGA ACCTGATGAC CTGTCAGAAT
 151  CCCACCCTTC CTTCTGTCTC CCATCGAAGT CCTCCAGGAA ATGCAGCTGT
 201  TTCAGTGACA GGGGGTGATT GCCATCTTCC AACTGAGGAG GAATTTGGGG
 251  TTTTGGTCCA GTCCATGAAG TGTGACACAG TCAGAATAAA AGGCGTCCTT
 301  CAAGGCCCTA CCACAGCCCC TCCTCTCATG ACCAGTGAAG GCAATGTAAC
 351  TGCAGAAGAC ACTGAGGAGG CAATTCGGGC TTTTGTGTAC GCTGTGGCTG
 401  CTGCTTCTGC TGCCGAAGCT TGGCATTGGA GACACCTCGT CCTCCTCTCA
 451  GGACAGATCC ATGAACCCAT CGGCAGCGGC GGGAATATAA TCAACACTAA
 501  TAAAGGAGGA AGGAGCTGTC AGAATCCTGC CCTTCCGTCT CCAGATCAAA
 551  GTCCTTCAGG AAATGCAACT ACTTCAGTGA CAAGAGATAA TTATCATCTT
 601  CTGACAGAGG AGGAATTTGG GGTTTGGTCC CAGTCCATGA AGTGGCACAG
 651  TCAGAATAAA AGTGGCGGGA GCGTCCCCGT GAGAGGCCCG ACCCAGGAGC
 701  CATGTTCTGA ATCTCAAATT TTGAAAGAAT CTTTTGTCCC ACCCACAACA
 751  CCCAAAGAAA ATAATAAACA GGAGAGGGAG GATGAAAATT GGCGTCTACC
 801  ACCCCCTCCA GTAGCAGAAA CACCTGTACC ATCTCCTTCA GTAACAGAAA
 851  TAGAGACCCC ACTGCAAAGA ATTCCGCGGA CTGCTACCAT AGCTGGAGAG
 901  CCCTTAGGAC ATTGCACTTT CACTATTTCT CCGGCATTCG TACATTCTGT
 951  GCTCAACAAA CGGAAGCGGC AGCTGGAGCT GCTGCTCCGG GAGGTGGAGT
1001  GGCCTGGCAG AGGGCACATG GCTGCCACCT GCTGCAAGCT CCAGGTAGAA
1051  GGGCAGGACA GAACCATGAG CCTAGCGGCA GCGCCGGTTC GCGAAGCTCC
1101  CCCTCCGCCA ACGGGCGCCT CCTCAGAGCC GTCCGTGCCC GCCCTGCCGG
1151  GAGCTGACCC GCAGCGCAGT GCAGAGTTGC TCCTGTTGGC GGTGACCAGG
1201  GAGGGACTGG AGCGGCGGAT CATCTCCAGG AAGCGGGCTG AGTAG
```

FEATURES:
Start codon: 1
Stop codon: 1243

Homologous proteins:
EST:
Blast hits:

	Score	E
CRA|3000001473333 /altid=gi|1186824 /dataset=dbest /taxon=9606 ...	159	2e-36

FIGURE 1

```
  1  MGPWSVVVLV CGMKQLGQAL QASVSLSIIT ENQGKRCPFC GAQNLMTCQN
 51  PTLPSVSHRS PPGNAAVSVT GGDCHLPTEE EFGVLVQSMK CDTVRIKGVL
101  QGPTTAPPLM TSEGNVTAED TEEAIRAFVY AVAAASAAEA WHWRHLVLLS
151  GQIHEPIGSG GNIINTNKGG RSCQNPALPS PDQSPSGNAT TSVTRDNYHL
201  LTEEEFGVWS QSMKWHSQNK SGGSVPVRGP TQEPCSESQI LKESFVPPTT
251  PKENNKQERE DENWRLPPPP VAETPVPSPS VTEIETPLQR IPRTATIAGE
301  PLGHCTFTIS PAFVHSVLNK RKRQLELLLR EVEWPGRGHM AATCCKLQVE
351  GQDRTMSLAA APVREAPPPP TGASSEPSVP ALPGADPQRS AELLLLAVTR
401  EGLERRIISR KRAE
```

FEATURES:
Functional domains and key regions:

BLAST Alignment to Top Hit:

Hmmer results:

Model	Description	Score	E-value	N
PF00947	Picornavirus core protein 2A	4.3	5.6	1

Parsed for domains:

Model	Domain	seq-f	seq-t		hmm-f	hmm-t		score	E-value
PF00947	1/1	197	205	..	1	9	[.	4.3	5.6

FIGURE 2

```
   1  AGCCAGACTA GGAGTGAGCC AGAAGAGGGG AAGGATGGTG GAGGCACAGG
  51  CTGCACTCTA CTGGTGCCCC AGACCCAGAC TGCATGCCCA GGCTGCAGTC
 101  CAAAGGATAC TCGGTGCGGG TCCCTGTCCC CCATAGCATC TTAGATCAGC
 151  TGCTGAGGCT GGAGCTTCTT CCATTCCTTG AGCATCAGGG GTGTGTATCA
 201  TTTCCAAGGG TTTTCAGACA ATCCCTGGTG ACCCCTGGCA GGGGGCGGTT
 251  ATCATGGCGA TCGGTCCATG GCCTTGCCTC CAAGCAGCAC CCAGCAATCC
 301  CCATGCCCAC CAATGCACTA AATGTTTGTG GTGGGCCTCT TTCTGGAAGC
 351  TCACCTTCTC CTCCTGTTTG CCCTCCATCT TCCCCAAACC AGTACTTCTG
 401  GCCATCCTCC TTGTCACCAC AATGGGAAAA CTGGGTCCTG GAGACTCAGA
 451  AACCACTGTG CAGGCCTCGA GTCTTCCCCT GTCCTGGCTA ACAGGGCATG
 501  GAATCAGAGA GAAAAGTCAT CTTCCACCTC CTGAAGGCTG CCAGCGTCAG
 551  GGCTTGGCAC ACTGAGGCTG ACAGGGGCCT TCTGAAGGCC AGAGGAGATG
 601  GCCCGGGACA TAAGGCTGAA GCAACCTGTC TGAGCCAAAG ATCTGTTTGT
 651  GTCCTCCTGA ATCTTAGTGG CCTTCTAAAG GCGGGTGTGA TCAGCCATGG
 701  GTATCAGAGA CACTGGAGTC CAGTAGCTGC TAGGTGGGAC ACGGGCACAA
 751  TTTCACTTGC AGACCAGCTG CACGGAGTGG ATAAAGAGAG AGTTCTGTGT
 801  GGGAATCTCC TTTGGTGGAT CATCAGGGAG GTGAAGTCTT TGTCATAGCC
 851  TCATATCCAG CTTGTGTGAT ACCAATTCCA GTGAAGCTGG AACAAGCTGG
 901  CACTGCTCAA ACAGGCCTAC CAAGACATCA TGTTTTTTTT TTTTTTTCC
 951  ACCAAACCTG GACCTGAATG GGGATGTGGA CACACATAGA GTCCAGAGGA
1001  TGGGACCCTG GTCAGTGGTG GTGCTGGTGT GCGGCATGAA GCAGCTGGGG
1051  CAGGCCCTCC AGGTGGGAGG AGGAGCCAGC CTCTCCTGTG GGGTCCTGGG
1101  CAAGTCATTT CCCTCCTTGA CGCTCTGTTT CCTCATCTGG AAAATCCAGG
1151  AAGCCTGTTG TGCAGTCCTC AGAGGGTCAT GATAAGGTGC AAAGGAGGAA
1201  GAAATTTTGA GAGTTCTTGT GCCTTCCTCT TCCTGAGAGG GATGATGGTG
1251  AGAACAGTGT GGATAACCAC ATGCAACTGA GTCCCCAAAA GGCCCTGTGA
1301  GGAAGGTGTT GTTCCCATCA TGCTTCCCAG ATGAGGAAAC AGTCTCAGGG
1351  AGGCCTGGCT GCATGCCCAA GGGTTACACA CACACTGAAT GTTGGAGCTG
1401  GGAGTAAATC TAGAGTCAGG GCTCACTGGA GGTGGTGGGA GCATTGCACC
1451  AGCTGCCTCA TGCAGTTATC CAAGATCTGA GGTCAGGGGT CTGGGGTATA
1501  ACTCGGTGAG AGAACCACAG TCTCACTGAC CCTGTTTCTG ACTCTGCTAG
1551  GTAGGGAACT TTTTAGAGTC CAGATGCTGC TGTTGCTGCT CCAGGGAGTT
1601  AAGGACAGGG CTTTTCCCTC CCAGCTGGAG TTCTCCACAT CAAGAGACCA
1651  GAGTGTCTTC CTCCTCTAGC CCTGCCCTCC TGTGGCCACA GTACTTGGAA
1701  TGCCTTCATT AACGTGACCA GAGAAAGAGA CTTAGAGAGC TGAGTTCACA
1751  GTGGATTGGG AAAAGAGTGA GGCTGGGGAC AACCTGACTT TGTGACTTAT
1801  TAAAATCCTA CCTTAGAGTT AACTAGAAAG AGCTTGGTGT CTTTGGAGGC
1851  CAGCTGTGGA AAGAGGAGGT AAAGATGTCT TGTAAGGAGG CTGAGGGTCC
1901  TTGGCTGCTC CAAGTGGGAG ACTTGGGTAG GGGCGTCAAG GAGACATGGG
1951  TTGGTGAGAG TTCAAAGGAC AGGCCTTATG CAGGAAGCTG AGAATTTGTG
2001  CTCATCACTG CTACCACCTT GGAGGAACAG GGAGCTGTAC CCTCTCTCGG
2051  CACACTTCTC ACTGACCTGA TGATGCTGGA CATTGCCATA GAGGATTATA
2101  ATGATGTGTG TGACACTGGG GCAGGCAGGC AGTTTGGGAA CCAAGATCCT
2151  GAAGCTTGTG AGAACAGAGT CCTGGAATAA GCCCTGAGAT TGCAGCACTC
2201  AGCAAACCTG CTTTCCTGAG AGCCTACCAG ATGCCCATGT GAAAGGGGGG
2251  TGTCAGTTCC ATCTTATCTA TGAGTGACAG AGGCTCCAGC AAAGACACCA
2301  GTCACTGTTC CTTGAGTAGT GAAGCTGCAG AGCTAAGTCT CAGCCTCTGC
2351  CTTAGGTGGT ACCCATTGGA GAAAACAGAA ATTGGGCCCC TCCTAAGACA
2401  GGAAGTTCCT AAATTGTTGA GTGCCTTTTT TGTTTCCTAG GCCTCAGTTT
2451  CCCTGTCTAT CATCACAGAG AATCAGGGCA AAAGGTGTCC CTTCTGTGGA
2501  GCCCAGAACC TGATGTAGGT CCAAGTCCTG TTTTATGCAC ATGCCTTGAC
2551  CCTGGCAGCC CTGGCGGTGG TGCAGCATGG GAAGTACAGG GGATGAGGGC
2601  TAGTCATGGG CCAGGGGGTC TTTCTGAGGG ATCTTGGCTG TCTACCTTCC
2651  AGGAAAATAT AATCAACACT AATAAAGGAG GAAGGAGAGC AGCTGGGGTC
2701  TCACTTTGAG GGAGGCTGGG GACGTGACAG TCAGACACCA CCCTGAAGAG
2751  GCCACTCGCT GGCTTCACCC TCTGCATCTT AAAGTTATTG GGAAGGTTTG
2801  ATACACAGAG GAGATCCATT CTAATGGAGG GTTTGATTAG GGGACTAGAA
2851  TCAACAATAA ATTCCTAGAT GAGGAACTGT TTATATCCAA CTCTGAGAAC
2901  AGGTTAGGGT TACATGGGAT TGGAAGAGAG GGTGGGGTCC CTTAAAAGAA
2951  AAGCCCCAGA AACTCACTGC TGCTCTATCC CTCCCCTATA AGTTCTCTTT
3001  GTTATCTTCC ACCCAGGACC TGTCAGAATC CCACCCTTCC TTCTGTCTCC
3051  CATCGAAGTC CTCCAGGAAA TGCAGCTGTT TCAGTGACAG GGGGTGATTG
3101  CCATCTTCCA ACTGAGGAGG AATTTGGGGT TTTGGTCCAG TCCATGAAGT
```

FIGURE 3, page 1 of 41

```
3151  GTGACACAGT CAGAATAAAA GGTGAGGGCC TAACAGATTA GCAGACGGTA
3201  GGAGAAGACT ATCTTGCAGC CAGCTTCAGA GAGCCTGTGG CCATGGCTCC
3251  CAGGTCAACA TTAGGCCCTG TTGCCTGGGA ACCCCTGGGC AGGCAGTGGG
3301  AAGGTTGAGG TGTGGCTCCT GGTAGCCTCA GAACTGCCAC TATTTCCTGA
3351  AGCTCCTACT TGTTCTGTCA GCTAAGCCCC CATCCCAGTA GGCCAGCAAC
3401  ACCCTCAAGA CCAAGAACAG GCCATGGTGA ATCTCAGGGC CACTAAGTAC
3451  CTGGGCTGGC AGGGGCAGAG TGCCTCAGGG CTCAGTGTTG TTTGGGCTGA
3501  GCATGGGCTT TGGGAGTCAG ACAGCTGCAC TGGGCTCCCA GCTGCACCAT
3551  GGCCAGCCCT GTGTATGGGG CACTGCTCTG TAACTTGAGA CACCATAGTC
3601  ATAAATATAA CACACCCTTC TAACTGCTTT TTCTTTTTTG TCTATTTTTC
3651  TCTATAATCC CCATGTACTA CTGATCTGTT TATTTAAATT AATAAACATG
3701  TTATACAGTG TATATTGTTC TTCCTCATGA TTTCTTTACT ATATTGTGTG
3751  ATTCCACTCA TATGAGGTTC TTATGGATGT CCCATTCACA GAAACACAAA
3801  GTAGAAGAGT AGTTAGTTCC CAGGGGCCAA AAGAAGGTAA ATGGGGGCTG
3851  TTTCTATCTT TTATTTTTAT TTTTGCAAAA TGAACAAAAT TCCCTATGAA
3901  TGTGGATGAT GGTTGCAGAA CAATCTGAGC ATGATTAATT CCTCTGACTT
3951  GCACGTTAGA AATTGTTAAA ATAGTTAATT TTATGTATAT TTTACCACAA
4001  TGTAAAAAAG GAATTTTTAA ATGAACAGAC TGTAGATACA TGCAACAGCA
4051  TAAATGAATA TCACAAATAT AATCTTGCAT TTAAAAATTG ATGTAAAAGT
4101  ATCCAAACTA TATAATTTCA TTTATACTAA ATCCAAAAAT CAAAACTGAC
4151  ATTCTTGCTT TCACTAATGG GAATTAGCTA GTTAAACTAA CACTCTCAGA
4201  GAGAAAAATG ATGAATCCTA GATAAAATAG TATATATCAT TATAAACACT
4251  TCTATATATA ATACATATAT GAGATATGTG TGTATAAGAA GTGAATGAGG
4301  ATGTCCCCTG TGCCCTCCTT AGGAGAGACA AGAATTGAAG TTATAATCCA
4351  GGCCAATTAG CACTCTCTTT AAAAATCAAC ACTCTTCAAA GGGACACAAC
4401  AGAATCCAGA GTCTCTATAA CTCTTGTATA CAGTCTCTTG TACACAATTT
4451  TCAAATTCGT GAGATGGGTG AAGACACATG AAAATGCAAT ACATACACAA
4501  GATAAAAGGC AGGCAGTAGA CATCTCCAAG ATATCCAAGA TGTAATCAGC
4551  AGACAAGAAT TTGAAGGCAG CTATTACAAG TATGCTAATG GAGGCAAAGG
4601  AAAAAATATA CTTATAAAGG AACAGATGTG GAACCTCAGC AGAGAAATAA
4651  AAAATAGCCA AATAGAAAAA TAAGACACAA AAAGAATAAT TTTGAGCTTA
4701  TCTATAGATC AGAAACAAAA CACACAACAA TAGAAATTAT TCAATCTGAA
4751  GATACAAGTA AAAAAAAAAG TTTAAGGAAA ATGAACCCAG CCTTACAGAC
4801  CTCTCATGGG GCACTCTGGG AGTTGTAGTC TCTTCTCCGG TTCCAAACGG
4851  TGGCTGTTGT GGCTGCAGGA TAACAGTCCC AGATTGAGAC AGGGCAGAGG
4901  CTGTGTGCAG CCCTACAGGA AGGGGCAGGG TGGTGTAGGC CTCTTCACTT
4951  ACCAAGATTT GCTGGCCATT GATTCCGTGC CAAACCCTTC CCAAGGGGAT
5001  TGAGTCAGGA GAGGATCTTG AGAGTCACTC AGGGTCTTCC CAGAGCATCT
5051  GTGCCTCCTC CAGCCCACGG AGCTGCCTGA TTTCCTAAGT GGCTGTGGGA
5101  ACTGGTCTGA AGTACCAGAC GCTGTCTACT GTGCTGCTGC CCTCTGTTCT
5151  ATCTAACCAA AGTGCAAGTT CAGCTGCCTT TGAAAGACAT CCACTGCCTG
5201  ACCTGGGGAT GCACGGGTTC AGAGCTTTGC AGGGAGTGAA CATGGGCTGT
5251  GGCTTCATGA AAATATCACC CTCCCCAACG CGTTTTTGCA GATCTGGACT
5301  TGGAGGCACG AAGGACGGTA ATCATTGGGT TACCAAGGTG TTACTAGGAG
5351  CAGAGGAGAA AACCGCAATT CCTAGCCATG TGTCTGGTGT GACATTTCGC
5401  CAACCCATTT AAGTGTGCAG ACCCCCAAAT ATCTACCTAA AGATTATGAT
5451  AGTTTAGGCA TTTTACATTT AAAATTATTG GCTTCATGTC CACTGAAGCC
5501  TGACTGGCCA GTGTCTCAAA GACACAGATG ATGATCTGAT CCCTCAGGAA
5551  CAGATGGTTC TCCAGCTTTG TTGGAGTGAC TTTCAAGGTA TGGAGCACTT
5601  ATATAAATTT GCCTAAGAGT AGGATTTGTG CTAAGTACCT GTTCACAATA
5651  ACATCAAGGT TGTTTTGATT TAAGGGTAGG GCTTACATAA GCAGTAGATT
5701  TCAATATATA ACATAGATTC TTGAAACCCC CCCAAAAAAC ATTAAAGGAA
5751  GTACCTATGT CATAATTTTA ATTTTTTATT TAGTAATTTA AAATCTTAAC
5801  GTCTTGTTTT GTTAGCTAAT CTTAAGTTTC TCACTAAAAA TTAGCATGAT
5851  TAAGCATGAA AATAATAGCT TTAAGACAGT TTTTACCCCA GAACCAGTGA
5901  TTGGATAATA GGGTTCCAGG CCCTCCCCTT CAGGTCCTGC GTGACAGAAT
5951  GTGAACCAAT TCATAGCCAA GCGAGGAGAG AGTGAAACGT TCCTAGGTGC
6001  AGCCCCTTTC AGGCAGGACT TACTCCTTAT GCTGAAACCT GGCCCTCACT
6051  ATGAGACATT TGCATTTAAC CTTGTATATA AGTTTATTTT TATTCATAAA
6101  TTATATATAT GCACATATAT GTATATATAC AGCTGGACAT GGTGAATCTC
6151  ACCTGTAATC CCAACACTTT GGTAGGCTGA GGGGCGAGGA GCTCTTGAAA
6201  CCAGGAGTTC GAGACCAGCC TTGGCAACAT AGTGTGAGCA CCCTCCGCCC
6251  CCCAACCTTT TCTACAAAAA AAAGAAAGAA AGAAAAAATA GCCAGGCATG
```

```
6301  GTGGAGCTTG TCTGTGGTCC CAGCTACTTG GGGGACTTAG GTGGGAGGGT
6351  CATTTAAGCC TGGGAGGTAG AGGCTGCAGT GAGCTGAGAT CAGGCCACTG
6401  CATGCACTCC AGTCTGAGTG ACAGAGCGAG ATCCTCTCTC TCTATCTCTT
6451  CCTCACTCTG TGTGTGTGTG TTGGGGAGAG GGGTGTGTGT TTGTGGGTGT
6501  GTGTGTGAGT GTGTATGTGT GTTTATTATT CAAAATGAAA ACAACATAAT
6551  GACAATTATT TTTATTTTTA TTTTTTTGAG ACAGAGTCTC ACTTTGTCAA
6601  CCAGGCTCCA GTGCAGTGGT GCGATCTCGG ATCATTGCAA CCTCCGACCC
6651  CGAGATTCAA GCAATTCATC TGCCTCAGCC TCCTGAGCAG CTGGGATTAC
6701  AGGTGCCCAC TACCTTGCCT GGCTAATTTT TGTATTTTTA GAAGAGGCAG
6751  GGTTTTGCCA TGTTCTCCAG GCTGGTTTTG AACTCCTGAG CTCAAGTAAT
6801  CCGCCCACCT TGGCCTCCCA ATGTGCTGGG ATCACAGGCA TGAGCTGCCA
6851  TGCTCGACCA ATGACAATTA TTTAAAAATT TTAGATTTTA CAATCTTTCT
6901  GGCCTTTTGG CTTTTGAGGC AGCCTGAGCT GTGAAAATAG GCAATCCTCT
6951  ATTGCAGCAA TGTGCAATAG AAGTAAAATG TGAGCCACGT GTGTCAAATA
7001  AAATTTTGTA GTAGCAACAT AAAAAAAGAA AAATGAGTGA AATTGATTTT
7051  AATAACAATA TATCGGAAGT ATTTTAACAT ATGATCAGAA TTAAATTATT
7101  TTATATATAT ATTTTGGGAA GCACACAATT CAGCTCACAG CAGCCTCTGT
7151  CAGGACATAA CGTTTACATT GAGATCCCAG TAACCTAGCA ACAGAGGGAA
7201  TGACATACAA AGATTCTGGG GAAGAAGATT CCACACAGAT GATTCTATTG
7251  AGCAATGGTC CTGAAATGGG AGTGAACAGG GTGAGTTTTA GGAACCAGAG
7301  ACCAGTGATG GCAAGGAAGA GGTATGAAAA GCAAGGAGAG GAGGTGAAAT
7351  CACAGAGATC CAGTGAGATA TATAAACTTG ATTGTGATTT AAGCAGTTTC
7401  TACCTTTTGG ATTTTGAGAA CAATACTGTT ATTAACATGA GTGTGCCAAT
7451  GTTTCTTGCA GGTCCTGCTT TGAACATTTA GATAGATATC CAGAAATGAG
7501  ATTGCCAGAT CGTATTAGAA TTCCATTTTT AGTATTCTGA GGAATATCTG
7551  TACTATTTTT CATAATGGCT GCATTATTAT TTTTTCCACC ACCAGTGTAC
7601  AGTGTTCCAA TTTCTCTACA TCCTTGAGAA CATTTGTTAT TATTTCTTGT
7651  TTGATAGTGG CCATCCTAAT GAGTGTGAGG TAATATCTCA TTGGGATTTT
7701  GCTTTTTATT TCTCTCAAGA TTTGTAGTTT TGAGCATCTT TCAAATTCCT
7751  CTTGGCCATT TGTATATCTG GTTTTTAAAA ACATACGTTG ATCATTTTGC
7801  CCATTTTTAA ATAGGGTTAT TTACTTTTTG TTGTTGAGTT TTAGAGGTTG
7851  TTTATACATT CTGGATATTA ACGTCTATCA AATATGTTAT CTGCAATTTT
7901  TTCTCATTTC TTAAATGACA TTTTTACTCC ACTTAATGTT TTCTTTGATG
7951  TCCAGAAAGC TTATTACACT TGATGTAGTC CCATTTTTCT GTTTTTATTC
8001  TTGTTACTTC TGCTTTTAAT GTCATGTTCA AAAAATTACC AGGACAAATG
8051  TCACAATTGT TTACCCTATA ATTTATTTTA AAAGTTTTAG AGCTATCTTA
8101  CTTACATTTA AGTATTTAAT TCATTTAAGA TATTTTTGTA TACAGTGCAA
8151  GTGAAAAGTT AAATTTCATT TTTTTTAATT TTGATATTCG GTTTTGTAAC
8201  ACTATTTGCT AAAGCGTCTG TTCTTCCCCT TTGTTCGGTC ATGGCAACTT
8251  GATTGAAGAT TATTTGCTGA TATTCATGAA GATTTATTTC TGGGTTCTCC
8301  ATTCTGTTTC ATCATCTATT TGTCTTTCTG TTTGAATTTC TACAGCTTTG
8351  TAATATATTT TGCAATAAAG TTGCAATCCA ACTTTGTTCT TCCCTACAGC
8401  TATTTTGGCT ACTCATTGTC CCTTGAGATC CCATATGCAT TTTAGGACTT
8451  AAAATAATAT TTCTCCAAAA AAGAAAACTC AGCTTTTGTG CCAGACATGC
8501  CTTGAGATTC CATATAAATT TTAGGACTTA AAATAATATT TCTCCAAAAA
8551  AGTAACATTG AGATTTTGAT ATAAAATACT TTTCTTTGAA TTTGTGCTTC
8601  AATCCAAGTA GTATTGACAT CTTAACAATA ATAAAATTTC TGATCCTTGA
8651  ACAAGAGGTC AAGAGTGTGC TGTTTTAAGT TTCATATATA TTTTGATTTG
8701  CCAGTTTCCT TCTGCTTGTG ATTTGTAGAT GAGGGCTTAT TATGTTGCCC
8751  AGGCTGGTCT CCAACTTTTG GCCTCAAGCT ATTCTCCGTC ATCAGCCTCC
8801  TGATGTATTT GGATTACATA GATAAGCCAC TGCACCTGGC CTCTTTATTG
8851  TTTTTCTTAC ATTTTTATGA TTTGAAGGTA ATTTTTGAAA AGATTTAAAA
8901  ATATGTATCT CCTTAGAAGG TTTTCATTTT TAATGTAGTC AAAAACACGT
8951  AAAATTGACC ATCTTAAATA TTTTAAGTAC ATAATTAAAT AATATTAAAT
9001  ATATTTGCAC TGTCATGCAA CATATCTCTA GAATGTTTTT GCTGCAAAAC
9051  TGAAACTCAA TATCTATGAA ACAACAACTA CCCCTTTATC TCCTCCCCTG
9101  AGGCTCTGAC TACTTTCTGT TTCTAGGAGT TTAACTACTT TAGATATCTT
9151  ATTTAACTGG AATCACACAG TGTCCTTTTG TGGCTGGTTT ATTTTATTTA
9201  CATAATGTCC TCAAGATTTA TGTTTAGTGT AAAAATAATC AGATCTCCTG
9251  CTTTTAAAAA ACTGAATAAT ATTCCATTGT TTGTATATTT CAAATTGTCT
9301  TTACCTACTC AATCACTGAG GGACGTTTGG GCTGCTTCCA CCTATTAGCT
9351  TTTGTGAGCA ATGCTGCAAT GCATATGGAT ATACAAATAA TTCTTCATTT
9401  GGCCATATAT ATGAGATTTT ATTTCTGTGC TCTGTTCTTT TCCGTTGGTC
```

FIGURE 3, page 3 of 41

```
 9451  TGTCTGTCTG CCTTTATGCC AGTACCAAAT GGTTTGGTTA CTGTAGCTTT
 9501  GTAATACATT ATAAAGTCAA GGAGTGTGAT GCCTCCAATA GCATTTCTTT
 9551  CTTTGAAGTT TGTTTGGTTC TCGATACTCA CTTTAGATTC CATATAAGTT
 9601  TTAGAATTTT TTTTTGTATT TCTTCAAAAT AATATGACAG TTAAAATAAG
 9651  ATGGAGATGG CATTAAATCT GTAGATCACT GTGTAATGTG GACATCTTCA
 9701  CAATATTGTC TTCCAACCCT TGAATAAGAG CATGCTCAAA AGTATGTTGT
 9751  TTAATTTCCA CATGTTTGTA GATTTTGCAG TATTTTTCTG CTATCAATTT
 9801  CTAATTTTAT TCCCTTTTAA TAAAAAATAA TAGTTTGTAA TATTTTAATC
 9851  TTATTTTTTG TATGTTGTAT GTTGACACCC TAAACCCCAG TACGTAAGAG
 9901  TGTGGCTATA TTTGAAGAAA GTGTCATCAC ACAGATAATT ATGTTAAAAT
 9951  GAGGTCCTTG GGGGCACAGG TGGGGTGGTG GAGAGAAGGT CACATTATAC
10001  AAAATTTCAG TTAGGAAGAA TAAGTGCAAG AGATCTATTG TACTTGGTGA
10051  CTACAGTTAA TGTATTGTGT TCTTGACTAA TACAGTAGAT TTCGAGTGTT
10101  CTCACAACAA AAACATGATG GGTATGTGAG GTAATGCATA TGCAAACTAG
10151  CTTGGGTTAA CCATTCCACA ATATGTGTGT ATTTCAAAAC AGTACCATAA
10201  ATGCAGACAA TTTTGTGTCA GTTACAATCA AAAAAGTTTT AAAATGAGGA
10251  CCTTAGGGTG GGTCCTAATC CAATCTAAGT GATGTCTCCA TGAAAGAGGA
10301  AATAAGGATA CAAATGTGCA CACAGAGAGA AATGGCCACA TGAGGACACA
10351  ATGAGAATGT GGCTACTTAC AAGCCTAGGA GAGAGGCCTC CGAGAAAACA
10401  CACCCTACCC ACACCTTGAT GTTGGACTTC ATCCTGTAGA CGAAGTCCTC
10451  CACCCTCCTT CATCAGGTGG AAGCCTTTGA TTCTGAATAT TCTCCAAATG
10501  CTGGAAGGTA CAAAAGTGAA GAGACAGCAC AGACCTCAGG GTGAAAAGTT
10551  TAAAGAGAAT AACATCTTTC CATTGCTGTG TCCTATCCCC TACACACACC
10601  TATTCCAGTC TTTATTGGTC TTTTGTGTTT TCGTGTCCTG GGTATAGTGT
10651  TAGTTGTAAA TCTGTGTTTA CATACAGGAT AACATAAAAC AAAGGTAAAC
10701  AATAAAATAA AAACAGACAG CAAAACTCAA CTAATAGTGT TTGGGCATGG
10751  TGACAGTGAA GACAGGAGAG TCACATAAAA ATGGAGGTGG AACTTTTGAG
10801  CTAAATCAAT GTCCTGTTGT GTTTCTGTGT TTCTACATCA GACTCTATAG
10851  TGGCAATTGT CAGGTTAGGT GTTTTTATCC TTGCCTGCTC AGTAAGTGCC
10901  AGAGGAGATT TTTCTAAACT GGGTGAGGAA CAGGTAGAGA AGTGTAAGTG
10951  AGACAAACTT CCCTGCCATT TGCCAAAGTG GCAGCACATA ATGTATCATG
11001  AGTGCCTCTA CCCTCTGATA TCCAAAATAT CAACTCTATA GATGATGTTC
11051  TGTTGGCTCC TAATTTAGAA GCATCTGCCT TCATACTCTT TCCTAGAATG
11101  GTAACATCTC TCTGTCAGTA GCTGATAAAA TTACTAAAAT CTCATAGTTC
11151  TGTTTTCCAG TTAAGGTTCA CTGCAACTAT ATGGGCAGAT TTACATAGCT
11201  CAATTCTTCC AGCTTGACAT TGTTTTCTTT TGAAGCTTCT GAGAGAGGGA
11251  GTCAGCCTCC ACCTAGAGGT GGCCCTTGGA GTTTTTGACA CAGAATCTCC
11301  CTGACACTAC TACTTACACT GATTTGAAAG TCAGTGTTGA GGTTGGCTTG
11351  CTACAATGCT CTTGTCAAAC TGAATCCTGC CACATCAAGG GCTTGGGGCT
11401  TCCCAGCTTA ATTTTCCAAT TTTGAAGTGA GAGAATTTGA GTTCTACAAG
11451  ACATCAGAAG ACCGCTTAGA ATATAACACA TTCTAAAAGT AAATCGGAAT
11501  GCCACAGGAA CATCTTAGTT GTAAAAGAAA AATTAGGTTC TTTGGTAAAA
11551  ATTTTATGCC CCTTGATATG GTTTTGCTGT GTCCCCACCC AAATGTCATC
11601  TCAAGTTTTA GCTCCCATAA TTCCCAAGTG TTGTGGGAGG GACCCAGTAG
11651  GAGATAATTG AATCATGGGA GTGGGTTTCC CCATACTGTT TTCGTGGTAG
11701  TGAATGAGTC TCATGAGATC TGATGGTTTC ATAAGGAGAA ACCCCTTTGC
11751  TTTACTCTCA TTCTCTCTTT TCTTGTTTGC CACCAAGTGA GACATGGCTT
11801  TTACCTTCTG CGATGATTGT GAGGCCCCCT CAGCCATGTG GAACTGTAAG
11851  TCCATTAAAC CTCTTTCTTT TGTAAATTTC TCAATCTTGA GTATGTCTTT
11901  ATCAGCAGTG TGAAAATGAA TGAATACACC TCCCTTCACT GTTTGAAGAA
11951  AAACTGCTGG CTCTATGAGG TCTTCAGTTC ACTGAATATT TTTTAAATGC
12001  CTAGCCCTAA CTCACTGACA AGTCAAGGAA GCTGCCACCT TATGGTGTTC
12051  ACTAGGCTAC TGTGGATAGC CCTCATTGCC AGGCACACAG AACCTGAAGC
12101  AGGGGTGGCT GCTCCACTTA ATGGTGAGAG CTCAGGGTTT TGGGTCATTG
12151  TACATGTTCT AACTGTTTGG TCTTCCACAT TGAAATTAAA GGCTATTAGA
12201  TTAAGGACAT CCTTTTTCAG ATTCAAATCA TGCAGAATTT CAGCTGCTGA
12251  TCTGTAAGGT CATTCATTTA AGAGTCCATG GTAAGGGTTG GTCCTCCGAA
12301  AATGCTAACC ACAGTTAAGC CAAAAACCAA GCCTGAACCC ATGTGTAATG
12351  CAGTGGTCAT CACTGCATGC CAGATTTGTT CTCCTTGGTG AAAGTTGACC
12401  TATTTGTCTC ATGGTTTAGA AGGCCCCAGA CCATTCCTGG CATAATAACT
12451  GCATTGGCCT TTGGCCCACT TCCTGAAGTG TTGGTCATGT CTCACCACTA
12501  TTAACACTTT TAGGATAATA ATCAGCTGAC TCCAGCCAAA GCATGTGTCC
12551  CTTGAAGCTC ATTCATGTAA TATTTTAGGC TTTTGAGACC AATTGCACTT
```

```
12601  AAATCACATT GTAACTTTTA TCACTAAATC TGTTAACATG GCCAAACTGT
12651  TTGATTTTAC TTCCTTTTTT CCTTTAGGTT CACACATGTT GGCTACGTAA
12701  TTTGTTGATT GAATGTGACT GTCCCCCCAG CACCCAAAAA TCTCTTCTTG
12751  ACTGACTCCA AGATGAAGAG AACATCATTG AGTTCTATAG GCATCCATTT
12801  TCAAAATTTG ATATTTCTCA CTGACCTTTC CTGGACAGGA TGTGTCTTTC
12851  TTTTCCTAAG CTGCAACCCC AGGCAAATCT TGTTTGTACA TTCTCTGAAT
12901  GGCAGTCCAA CCCAAAGTGG GGTTGTGGTT TTCAACTTAG TTGTGGTGAA
12951  CATTTACATA GTTCTTGCAT CCATATCATC TGGCTCTACT AGAAAAAAAA
13001  ATTACTAGAA GACAGTATGG GTGACTAGAA TAGAAAAAAT GACTACTTAA
13051  AATTATAGGT ACTGGCATAG CACACATAAA TTCTGTGGCT GTAGAGGAAA
13101  GACAAAAACC CAAAACCTAA AAGTGAACAA ATTAGACCCA GGAACTACAG
13151  AAAAGGAGGG AAGTTAATAT CTTTCTGTCT TTCTGAATTA TCCCTAGTGT
13201  GGCCCAGAGA AGAGTTGGGG AGTCCTGTTG GGCCAGAATG AGTAGTAGTA
13251  TTAGTTAGAC CAAATGGTAG AACAGGATGG GTGGGTTCAG TTTCTCTAAG
13301  GAATTTCATC CTTTTGTGAT GTAGACATGG AATCCCAGTG AGCATCCTGC
13351  ACTCTCAGTG TCCAACTGTC TGAAACCACA GTTGTTTCTG AAGACTGAGA
13401  ACAATTGCTT TTTGCTGAGG AGACAGCACT CCACAGCTGA GGTAAGGTGG
13451  ATTCTATAGA GCATTAATTC CATTTTCTCC TTCTTCTGTG AGCTGGGTTT
13501  CTTTCTTCAG TTTTTTATCT AGAGATGAAA TTACATTGTC AATATTTGTG
13551  TAAAATAGAG ATATAAGTCT GGCAAGTAGA CGCTTAGATG CAACCCTCAG
13601  CAGAAATTCT GGTCTGTTTT TCTTCTCACT TTTATTCAAT TTTTGTCACA
13651  GAAAGATGCC CTGACATGTG GTCTCTCAAG AGTGATTAGA ACATTGTACT
13701  TCTAGAGGAG TTAGTTGATG GAACATGGCA GAGCTTAGAA TGAGGATGGA
13751  AGCCTCTGTC CAAATCTCCC AGGCTATCTA TGTGTGGGAA GTGGAGTCAG
13801  GAAAGACCCT AGAGCTTTGG AAGTGATCAT TAAGAGAAAA CAAAATCCCT
13851  GTAAATTAGA GTACAAGGGA ACATATTCAT TAGCCACCTT TAGAGTCAGA
13901  TGCTCAGGGC TGAGCTGGTG TGGGGGAGTG TTCAGACCTG TGCATGCCTG
13951  GGAAAGCCTC CGATCATGTA TGAATGGTGA GGCTTCTGGG TGCATGAGAT
14001  TGAGTGCCCT TCCTTGGCAG AGCACCACTG AGTGAACAAT TGATTTAGGA
14051  TCAAGCATAT GGAGATCAAC TTTATTTTGA ATATCTCAAA GACAGAAACT
14101  GAAAAGATTT TTTTGCACTT TGAAATTGAG TAAGGGTGTC AGAAACTTCA
14151  GTAAAAAAGT CACAGGAGGA AACCCAGAAG TCTTATTCAT CCCCAGAAAC
14201  CACCAAAATC CTGATCCAAA CTGTAAGAAT TCACCTCGTA AAAATTTGAT
14251  TTTTGAATAG GAACAGCTCC GGTCTACAGC TCCCAGTGTG AGCGATGCAG
14301  AAGACGGTTG ATTTCTGCAT TTCCATCTGA GGTACCAGGT TCATCTCACT
14351  ATGGAGTGCC AGACAGTGGG CGCACGTCAG TGGGTGTGCG CACCGTGCGC
14401  GAGTGAAGCA CGGTGAGGCA TTGCCTCACT CACGAAGTGC AAGAGGTCAG
14451  GGAGTTCCCT TTCCTATTCA AAGAAAGGCA TGACAGATGG CACCTGGAAA
14501  ATCGGATCAC TCCCACCCGA ATACTGCGCT TTTCCGACGG GCTTAAAAAA
14551  TGGTGCACCA GGAGATTATA TCCTGCACAT GGCTTATAGG GTCCTACGCC
14601  CACAAAGTCT CACTGATTGC TAGCACAGCA GTCTGAGATC AAACTGCAAG
14651  GTGGCAGCGA GGCTGGGGGA GGGGCGCCCG CCATTGCCCA GGCTTGCTTA
14701  TGTAAACAAA GCAGCCGGGA AGCTCGAACT GGGGTGGAGC CCACCATAGC
14751  TCCAGGAGGC CTGCCTCTGT AGGCTCCACC TCTAGGGGCA GGGCACAGAC
14801  AAACAAAAAG ACAGCAGTAA CCTCTGCAGA CTTAAATGTC CCTGTCTGAC
14851  AGCTTTGAAG AGAGAAGTGG TTCTCCCAGC AAGCAGCTGG AGATCTGAGA
14901  ACAGGCAGAC TGCCTCCTCA AGTGGGTCCC TGACCCCTGA CTCCCGAGCA
14951  GCCTAACTGG GAGGCACCCC CGAGCAGGGG CAGCCTAACA CTTCACACAG
15001  CTGGGTACTC CAACAGACCT GCAGCTGAGG GTCCTGTCTG TTAGAAGGAA
15051  AACTAACAAA CAGAAAGGAC ATCCACAACA AAAACCCATC TGTACATCAC
15101  CATCATCAAA GACCAAAAGT AGATAAAACC ACAAAGATGG GGAAAAAACA
15151  GAGCAGAAAA ACTGGAAACT CTAAACAGCA GAGTGCCTCT CCTCCTCCAA
15201  AGGAACGCAG TTCCTCACCA GCAATGGAAC AAAGCTGGAC AGAGAATGAC
15251  TTTGACAAGC TGAGAGAAGA AGGCTTCAGG TGATCAAATT ACTCCAAGCT
15301  ACGGGAGGAT ATTCAAACCA AAGGCAAAGA AGTTGAAAAC TTTGAAAAAA
15351  ATTTCGAAGA ATGTATAACT AGAATAACCA ATACAGAGAA GTGCTTAAAG
15401  GAGCTGATGG AGCTGAAAAC CAAGGCTAGA GAACTACGTG AAGAATGCAG
15451  AAGCCTCAGG AGCCGATGCG ATCAACCGGA AGAAAGGGTA TCAGCGATGG
15501  AAGATGAAAT GAATGAAATG AAGCGAGAAG GGAAGTTTAG AGAAAACAGA
15551  ATAAAAAGAA ATGAGCAAAG CCTCCAAGAA ATATGGGACT ATGTGAAAAG
15601  AGCAAATCTA CGTCTGATTG ATGTACCCGA AAGTGACGGG GAGAATGGAA
15651  CCAAGTTGGA AAACACTCTG CAGGATATTA TCCAGGAGAA CTTCCCCAAT
15701  CTAGCAAGGC AGGCCAACAT TCAGATTCAG GAAATACAGA GAATGCCACA
```

```
15751  AAGATACTCC TTGAGAAGAG CAACTCCAAG ACACATAATT GTCAGATTCA
15801  CCAAAGTGGA AATGAAGGAA AAAATGTTAA GGGCAGCCAG AGAGAAAAGC
15851  CAGGTTACAC TCAAAGGGAA GCCCATCAGA CTAACAGCAG ATCTCTCGGC
15901  AGAAACTCTA CAAGCCAGAA GAGAGTGGGG GCCAATATTC AACATTATTA
15951  AAGAAAAGAA TTTTCAGCCC AGAATTTCAT ATCCAGCCAA ACTAAGCTTC
16001  ATAAGTGAAG GAGAAATAAA ATACTTTACA GACAAGCAAA TTCTGAGAGA
16051  TTTTGTCACC ACCAGGCCTG CCCTAAAAGA GCTCCTGAAG GAAGCGCTAA
16101  ACATGGAAAG GAACAACCAT TACCAGCCAC TGCAAAATCA TGCCAAAATG
16151  TAAAGACGAT CGAGACTAGG AAGAAACTGC ATCAACTAAT GAGCAAAATA
16201  ACCAGCTAAC ATCAAAATGA CAGGATCAAA TTCACACATA ACAATATTAA
16251  CTTTAAATGT AAATGGACTA AATGCTCCAA TTAAAAGACA CAGACTGGCA
16301  AATTGGATAA AGAGTCAAGA CCCATCAGTG TGCTGTATTC AGGAAACCCA
16351  TCTCACGTGC AGAGACACAC ATAGGCTCAA AATAAAAGGA TGGAGGAAGA
16401  TCTACCAAGC AAATGGAAAA CAAAAAAAGG CAGGGGTTGC AATCCTAGTC
16451  TCTGATAAAA CAGACTTTAA ACCAACAAAG ATCAAAAGAG ACAAAGAAGG
16501  CCATTACATA ATGGTAAAGG GATCAATTCA ACAAGAAGAG CTAACTATCC
16551  TAAATATATA TGCACCCAAT ACAGGAGCAC CCAGATTCAT AAAGCAAATT
16601  CTTAGTGACC TACAAAGAGA CTTAGACTCC CACACAATAA TAATGGGAGA
16651  CTTTAACACC CCACTGTCAA CATTAGACAG ATCAACGAGA CAGAAAGTTA
16701  ACAAGGATAC CCAGAAATTG AACTCAGCTC TGCACCAAGC AGACCTAATA
16751  GACATCTACA GAACTCTCCA CCCCAAATCA ACAGAATATA CATTTTTTTC
16801  AGCACCACAC CACACCTATT CCAAAATTGA CCACATACTT GGAAGTAAAG
16851  CTCTCCTCAG CAAATGTAAA AGAACAGAAA TTATAACAAC CTGTCTCTCA
16901  GACCACAGTG CAATCAAACT AGAACTCAGG ATTAAGAATC TCACTCGAAA
16951  CCGCTCAACT ACATGGAAAC TGAACAACCT GCTCCTGAAT GACTACTGCG
17001  TACAAAACGA AATGAAGGCA GAAATAAAGA TGTTCTTTGA AACCAACGAG
17051  AACAAAGACA CAACATACCA GAATCTCTGG GACGCATTCA AACCTGTGTG
17101  TAGAGGGAAA TTTATAGCAC TAAATGCCCA CAAGAGAAAC AGGAAAGATC
17151  CAAGATTGAC ACCCTAACAT TACAATTAAA AGAAGTAGAA AAGCAAGAGC
17201  AAACACATTC AAAAGCTAGC AGAAGGCAAG AAATAACTAA AATCAGAGCA
17251  GAACTGAAGG AAATAGAGAC ACAAAAAACC CTTCAAAAAA TTAACGAATC
17301  CAGGAGCTGG TTTTTTGAAA GGATCAACAA AATTGATAGA CTGCTAGCAA
17351  AACTATTAAA TAAGAAAAGA GAGAAGAATC AAATAAACGC AATAAAAAAT
17401  GATAAAGGTG ATATCACCAC TGATCCCACA GAAATACAAT CTACCATCAG
17451  AGAATACTAC AAACACCTCT ACACAAATAA ACTAGAAAAT CTAGAAGAAA
17501  TGGATAAATT CCTCGACATA TACACTCTCC CAAGACTAAA CCAGGAAGAA
17551  GTTGAATCTC TGAATAGACC AATAACTGGA GCTGAAATTG TGGCAATAAT
17601  CAATAGCTTC AACCAAAAAG AGTCCAGGAC CAGATGGATT CACAGCCGAA
17651  TTCTACCAGA GGTACAAGGA GGAAATGGTA CCATTCCTTC TGAAACTATT
17701  CCAATCAATA GAAAAAGAGG GAATCCTTCC TAACTCATTT TATGAGGCCA
17751  GCATCATCCT GATACCAAAA CTGGGCAGAG ACACAACAAA AAAAGAGAAT
17801  TTTAGACCAA TATCCTTGAT GAACATTGAT GCACAAATCC TCAATAAAAT
17851  ACTGGCAAAC CGAATCCAGC AGCACATCAG AAAGATTATC CACCATTAGC
17901  AAGTGGGCTT CACCCATGGG ATGCAAGGCA GGTTCAATAT ATGCAAATCA
17951  ATAAATGTAA TCCAGCATAT AAACAGAACC AAAGACAAAA ACCACATCCT
18001  TATCTCAATA GATGCAGAAA AGGCCTTTGA CAAAATTCAA CAACCCTTCA
18051  TGCTAAAAAC TCTCAATAAA TTAGGTATTG GTGGGATGTA TCTCAAAATA
18101  ATAAGAGCTA TCTATGACAA ACCCACAGCC AATATCTTAC TGAATGGGCA
18151  AAAATTGGAA GCATTCCCTT TGAAAACGGG CACAAGACAG GGATGCCCTC
18201  TCTCACCACT CCTATTCAAC ATAGTGTTGG AAGTTCTGGC CAGGGCAATT
18251  AGGCAGGAGA AGGAAATAAA GGGTATTCAA TTAGGAAAAG AGGAAGTCAA
18301  ATTGTCCCTG TTTGCAGACG ACATGATTGT ATATCTAGAA AACCCCATTG
18351  TCTCAGCCCA AAATCTCCTT AAGCTTATAA GCAACTTCAG CAAAGTCTCA
18401  GGATACAAAA TCAATGTACA AAAATCACAA GCATTCTTAG ACACCAATAA
18451  CAGACAAACA GAGAGCCAAA TCATGAGTGA ACTCCCATTC ACAATTGCTT
18501  CAAAGAGAAT CAAATACCTA GGAATCCAAC TTACAAGGGA TGGGAAGGAC
18551  CTCTTCAAGG AGAACTACAA ACCACTGCTC AAGGAAATAA TAGAGGTTAA
18601  ATGGAAGAAC ATTCCATGCT CATGGGTAGG AAGAATCAAT ATCTTGAAAA
18651  TGGCCATACT GCGCAAGGTA ATTTACAGAT TCAATGCCAT CCCCATCAAG
18701  CTACCAATGA CTTCTTCACA GAATTGGAAA AAACTACTTT AAAGTGCATA
18751  TGGAACCAAA AAAGAGCCCG CATCGCCAAG TCAATCCTAA GCCAAAAGAA
18801  CAAAGCTGGA GGCATCATGC TACCTGACTT CAAACTATAC CACAAGGCTA
18851  CAGTAACCAA AACAGCATGG TACTGGTACC AAAACAGAGA TATAGATCAA
```

FIGURE 3, page 6 of 41

```
18901  TGGAACAGAA CAGAGCCCTC AGAAATAACG CTGCATATCT ACAACTATGT
18951  GATCTTTGAC AAACCTGAGA AAAACAAGCA ATGGGGAAAG GATTCCCTAT
19001  TTAATAAATG GTGCTGGGAA AACTGGCTAG CCATATGTAG AAAGCTGAAA
19051  CTGGATCCCT TCCTCACACC TTATACAAAA ATTAATTCAA GATGAATTAA
19101  AGACTTAAAC GTTAGACCTA AAACCATAAA AACCCTAGAA GAAAACCTAG
19151  GCATTACCAT TCAGGACATA GACATGGACA AGGACTTCAT GTCTAAAACA
19201  CCAAAAACAA TGGCAACGAA AGCCAAAATT GACAAATGAG ATCTAATTAA
19251  ACTAAAGAGC TTCTGCACAG CAAAAGAAAC TACCATCAGA GTGAACAGGC
19301  AACCTACAAA ATGGGAGAAA ATTTTCGCAA CCTACTCATC TGACAAAGGG
19351  CTAATATCCA GAATCTACAA TGAACGCAAA CAAATTTAGA AGAAAAAAAC
19401  GAACAACCCC ATCAAAAAGT GGGCGAAGGA TATGAACAGA CACTTCTCAA
19451  AAGAAGACAT TTATGCAGCC AAAAAACACA TGAAAAAATG CTCACCATCA
19501  CTGGCCATCA GAGGAATGCA AATCAAAACC ACAATGAGAT ACCATATGAC
19551  ACCAGTTAGA ATGGCAATCA TTAAAAAGTC AGGAAACAAC AGGTGCTGGA
19601  GAGGATGTGG AGAAATAGGA ACATTTTTAC ACTGTTGGTG GGACTGTAAA
19651  CTAGTTCAAC CATTGTGGAA GTCAGTGTGG CGATTCCTCA GGGATCTAGA
19701  ACTAGAAATA CCATTTGACC CAGCCATCCC ATTACTGGGT ATATACCCAA
19751  AGGACTATAA ATCATGCTGC TATAAAGACA CATGCACACG TATGTTTATT
19801  GCGGCATTAT TCACAATAGC AAAGACTTGG AACCAACCCA AATGTCCAAC
19851  AATGATAGAC TGGATTAAGA AAATGTGGCA CATATACACC ATGGAATACT
19901  ATGCAGCCAT AAAAAATGAT GAGTTCATGT CCTTTGTAGG GACATGGATG
19951  AAATTGGAAA TCATCATTCT CAGTAAACTA TCGCAAGAAC AAAAAACCAA
20001  ACACTGCATA TTCTCACTCA TAGGTGGGAA TTGAACAATG AGAACACATG
20051  GACACAGGAA GGGGAACATC ACACTCTGGG GACTGTTGTG GGGTGGGGGG
20101  AGGGGGGAGG GATAGCATTG GGAGATATAC CTAATGCTAG ATGACGAGTT
20151  AGTGGGTGCA GCGCACCAGC ATGGCACATG TATACATATG TAACTAACCT
20201  GCACATTGTG CACATGTACC CTAAAACTTA AAGTATAATA ATAATAAATA
20251  AATAAATTTA TAAAAAGAAA ATTGTTGTTT AAAATAAGTA AAAAAAATTG
20301  ATTTTTTTCA CCATATAGAT TTATCTTTCA TTTGACCTTT ATTTAATTAC
20351  AAGTTTTAGT TAATACATTT TATTTTACCT TTATGATAGA AATATCAGAT
20401  TCTTAAACTC AAAGCATTAA TGATGCCTAC CCAAGTGTAG TTTTTATTGC
20451  CAATTAACTT TCTACTATGA TGCAAATTTG TGTGTCCTTC TAAAAACTCA
20501  TTTGTAAAAA TTTAGTCCCT GATGTGATAG TTTTAAAACA TGTAGCCTTT
20551  TTGGAAGTGA CTAACTCAGG AGGGCTTCAT CCTCATGAAT GTAATTAATA
20601  CCCTGTAATA GAGGTTGAAG GGAGCACCCT TGTCCCTTCT GCCATTGAAG
20651  ACACAGCAAC AAGGCATCAT TTATGAGAAA TGGGACCCTC CCCAGACACT
20701  AAATTTGCTG GTGCTTTGAT CTTGAACTTT CCAGCTTCCA GAACTGTGAC
20751  CAACGCATTT CTGTTATTTA TACATGACCC AGTCTAATGT ATTTTGTTTT
20801  AGCAATCTGA ACAAATGAAG ACACTTTCTG ATGCACTGTG GTTTATTTTT
20851  GAATTTATAG TTCCACTGAG CTATCTATAT ATTCAAAAAT CAACATGTCT
20901  CACAGGGTGG GACAGCCACT CTAATTATTT TTTCAGAGTT TTCTTAGCTG
20951  TTCTTATTTG TGTTTTCATC TATAGGAGTT CTCCAATACA ATGCCTGCTT
21001  CCACTACCAA ATGGTATCTA TATTGGGACA AAATTAAATT TATTAATTAC
21051  TGCTAAAAAA ATTGATGATT GAGTAATAAT GAGTTTTTCT GCTTTAGAAC
21101  ATGATGTGAT TTTCTATTTG TTTATGCTGA CTTTCCTATA TTTCGAAGAC
21151  TTTTTATGTT CTTTGTCATA CATTTTACAA ATTCTTGTTA GATCTATGTA
21201  TAGCTAGTTC ATTTTATTTT GTCCTGTTGA AAAGTAAACT GTAGCACAAT
21251  ACAAATTATA CAAGTTTTCT TGGCAGGGAA TGATGATGCA TATCTGTGGT
21301  TCCAGCACTT TGAGAGGCCA AAGTGAGAGG ACTACTTGAC AAGCCAAAGT
21351  GAGAGGGGTT GAGGCCAGCC TGGGAAACAT AGTAAGACCC TGGCTCTACA
21401  AAACAGAAGA AGAAATTAGC TGGGTACAGT GGTGCATGCC TGTATTTCCA
21451  GCTACTCATG AGGCTAAGGC AGGAGGATGG CTTAAGCTCA GGATGTAAGG
21501  CTGCAGAGAG CTTTGATTCT ATCTATGCAC TCTAGCACGG GCGACAGAGC
21551  AAGAACCTGT TTCAAAAAAC ATTTTTTTCT TGAGTAACAA GCAACTAATG
21601  AATTGTGGAA CACCGGACCA AAAGAGGTTT ACCATTTTAG TGGCAAAGTG
21651  TCAGAGGTAT TGAAGAAATG CGGGAGCAAA ATATTAAAAT TATTTGATTC
21701  ATTGCAGGTA TAAAATTGTC TTATTTGGTT TACCTTATAG ACATATATTA
21751  CTATAAGCTT TTTGAGTATT TCTGATAACT TAAGCTTAAA TTCTGTTGTT
21801  GTTGTTTTTC TAATACAGGC ATTCACAAAA AATAGCTCGC ATTATGTTTT
21851  GCTTCTTTGC AAATCAACAA GATGATGTCA CTGGGGAGGT CTAACTGATT
21901  CTGTCTGCTC AGGAATGTTT CCAAGGCTTG GTCTCCTTTT TAATTTACTT
21951  TATAAATCAT TTTGTAATTT TTATCTCCCA CACATTCTGT ATCTTGTTAA
22001  TTACTTTTAT ACATATGTAA ATATATTTTA TGTATGATTT CTAAGTATAT
```

```
22051  CACTCATGAA AATTGACTGT GACACAGGCA TACCTTATCT TATAGCACCT
22101  TATTGTGCTT CACAGATGTT GGATTTCTTA CAAATTGAAG GGTTTTGGCA
22151  ACCCTATATT GAATGACTCT ATTAATACTA TTTTTCCAAC ATCATGTCAT
22201  CTTTGTGTGT GTGTATACCT CTGTCAGCAT TTTTTAGCAA TAAAGTACAT
22251  TTTTTATTAA GGTATGTATA TTTTTAATAT AAACATGGGC TATTTCAATT
22301  TTATTCTACA GTATAGTGTA AACATAACTT TTATATGTTC TGGGAAGCAA
22351  GAAAAATTGT GTGACTCACA TTATTGCAAT ATTTGTTTTA TTGCACTGGT
22401  CTAGAGCAAA ACCCATATAT CTCAAAGGTA CATCTATATA TTTTTATTGA
22451  ATTGGCCCCA TTAACTGAAA CGATAAATTG CTATTCTTTG ACTAAACAAG
22501  AGCTGTGGAG TGTGGAGAGG TCAGTGTGAA AATGAAGTAG AAGTGAATAT
22551  AAACAGTCTT CTAACTCACA CAGGTAGTAA TACAAATTAA TTATGGATGA
22601  AATATAAAAT TATCTAAAAT TGAATATTCT CAGAAACATT AATGTTTATA
22651  TCATTATGTA TATGGACATT AAGAACATAA TAAAAATAAA TTTAGATGAT
22701  GAAGGCTTTC ATAATCTCAA TGTAAAATGC AAGAATAAAA AATATAAAAT
22751  TTATCAATGT ATAAAATAAA TGACAATATT TATATTATAT TTAATATCCT
22801  GCAATACAGT ACAATCAACT GTAATGTATA AAATAGAAGA AAGTTTAAAT
22851  ATGGAAAATT TAAAAGGAGG CAAATATGAG TCAGGTAAAG AATAAATAAG
22901  TATGATCAAT TATATTTAAG ATACACTGAC CTGAGTTTTA TGGAAAAATA
22951  TTAATGGTAA ACACCTTGTG GAGTGATTTC ACCATAATTT TATTAAACTG
23001  TAAAAATATA TTAACATTTT CCCACAGGGA GTACAATTAA AGATTTGTGG
23051  ATATTAGTCC TCCATGGGCT CAGCGTGGGA AAGTTAGAGG CTACAGCCTT
23101  TTCTGAATTA AAAGAGAAAC ATATAACTTT TTATTTATTT TTCTACAATT
23151  TAAAATTTGC AAAGGCATAT GAATGATTAC ATCCTAATAT TTGTCTGATT
23201  ATATAGAAAT GCATGACTGT CACCAGACAT CTGAAAGACA TCAAATGTCT
23251  AACAAGAAAC ATAAAATTTG TTTATAATCT TAGCCCTCTG TGAAATGCAG
23301  GGTTCACCAT TTTGAGTATA TTGTTCAAGC TACTTTCCTA TAGCAGGTTC
23351  TTGCTCTCAT TCCAAGACCT GCAATTTCTG CTGAATGGGG TCAGGGTGCA
23401  CCCAGCTCAA GCCTCATCTG ACCCACTGAC AGGCTCAGTT ATCTCCTGCC
23451  CAGGCAAGGG ATGGGCTTCT CTATCCAGGG CTGGATCCCC AGGGCCAGGC
23501  AATGTGGCTG AAACAAGCCA GTTCTTAGCA GGGAAGACAT AACCTACCTG
23551  GGTGGCTATA AAATAGAAGG TCTGCACCTG AGCACATAGA GGCCGCCAGA
23601  ACCGGGCAAG CTGAGTGAGC TGCTCCCAGG TCAGTGGAGA ATGGACCTGC
23651  TGCACCGATA CCCAGTATAG GTCTTGATAA ATGGCCTTGA CACAGCTTGT
23701  AAAGTCACCA AGCTTTTCTG AAATGACAGC CATTGAACTC CTAGGGTCTG
23751  AGACCTGTGC TGCTTGGTGC ACCCAGTGTG AGTCATGAAA GGCCCTCTGT
23801  GGTGGGCATC ACAGGTCTCC TTGAGTTTAT TGCTGTGCAA AGTGGAGGAC
23851  TTTAGTTTCT TTTTCAACAT CAAGCTGTGC TCCTCTCCTG GACAGATCCC
23901  CGCAAAAGAA GCATGTGAGT GAGATACTCG CCAGCACAGT TCCCACCGAC
23951  CCTCATTTCC AAAACCTCCC ATGCACCTTC AGGTGAACAT TTGAATCTTC
24001  CCCTCCATCC TGACCATATT AGTACTTAAA TGAAGTGAAA TTTAACACCT
24051  TCTGAGTCCC CAAATATTCT CTGAGTGCCA GGATCTCAAA AATTTTTTCA
24101  GTCACCTTGC TGCTCAAACC CTCTATAAAA GTCAACTGCT TATTTTCCCT
24151  TTGGGGGAAA AAGGCAAAAT TCTACCAGCT CTGTCTTGGC AGCTGTCCTT
24201  GGAACCGATT TTCCTTTTCT TGGAGTTTCC CTCATGTGAG CTCGACTCTG
24251  GTTCTGTTGA TAAAATAAGA GTTTGAGTAA GTGTCTCCAA CCAAACACCC
24301  TAGAAGCCTT AGTTCATCCT GGACACAAAG GAGCTGAAGT AGCTATCAAA
24351  CCCAGCTCTC CTCTGTTCTC CAGAATCCAT GTCTATATGG CCCTGGCTGC
24401  CAAAGAGCTC CCAGTTTCCT TGCCAGGGGA GACTGTGTTG CAGCCCTTTC
24451  TCTTTTTACC TTGAAAGAGT CAAATTTTAC CTAATCTAGC AGTGCTGTTT
24501  CTAGCTTTGG GCTTAGTTTT CTCAGAATTC TTCTCTTCAT TTAGATTGGG
24551  CTCTGATCCT AGTGCAACAT GGAATTTAGG TGACTCACCT CTCTCAGACA
24601  CAGAGTCTCA TAGTCTATCT CTGACAAATA TTTGTGGATC AGTCCTTTAA
24651  GTGAAGCTCT TCTGCCAGTG TCATAAGTGA AGATGTTTCT CAAAGTCTCC
24701  CCAGGGCTCT AAGCCATCTT CCATCCCCAA TTTCAAAATA AAACCCTACC
24751  CAGAGACACA CAGCTCAGTA TCCCTGATTC CAACACTCCT TCCAGCCTCC
24801  ATAGGAACAG CCCAAGGCAT TACCCTGTCT TTGCCTGTGC TTCTCACTGG
24851  AATGGGAGGA GGGGGTCTCG GCTTTTTGTT TGAATTGTCT CTTCTTATCT
24901  GAGCCCTTTT CTGTAAAGGA GATCTGTTGG AAAGAAGGCT GGTCAGTGGG
24951  GCATTGGATG GAGGAGCAGT GGAGAATTAG GGTATTCATT TCCCTTTCCC
25001  TTGTTTAAGC TCAAGTGAAA GGTGCTCTCT CTTACACATC CAGATTCAGA
25051  CTGTGTGTTA TTCTGCTGGA CCTGCCTTTT ACATGTCCTT GGCTGGACGT
25101  GGCATACTCC TATTGACTGA ACAGTTTCCT TTTTCTTGCC ACTCATTAGG
25151  GCCTCATAAG AAAAGTTTCT TGCTGTAGTT TTACTAGGGA CCTAGACACA
```

FIGURE 3, page 8 of 41

```
25201  GTTAAAGAGG GACATTTTCT GGGTCTTGTC ATAGTGTAAA AAAACCCTAA
25251  ACAAACAAAA AACAAAGCCA GGGGCACAGG AATCAGAAAA TAGGGGAATC
25301  ATTTTTCTAA TTTCTGTCCC AATTCCACCT GGAAGTATTT ATGATACTGT
25351  TTATTTTTCA ACATGCAGAA TTAAACATAC CTATATTGAA ATGTGTCATA
25401  CATTTGGCAA AGGAAGAGAA TTACACATAG TGTTAAAATC ATGTACATAG
25451  ATATTATAAT TTTTCAAATG CTTGGAAATG TCAAATTAAA ATTATGGTTG
25501  ATTGTATTAA ATAGATACAT ATATGATAAC ATAAAAATAT GAAGAAAAAG
25551  TAAATACCAA ATAAAATGGC TCAAATAATT AGAGATTAGA CAATTAATTA
25601  GACAATAATT AGATCAAATA CAATCAACTC GAATTTATTT AATTAGAGCA
25651  GATGCTAACT TAATCAGCGC CTGATTCCTG AGGTAGCAAA AAGTCTAGGT
25701  GGAGAGAGAA ACTTACCCCT TTTCTTACCC TTCCTCGGTC ATCCTGGGAG
25751  CTCCACTTTC CTCTGTAGAA TTTATTCAGC CTCCTTAGTA AACATGGACT
25801  TGGTCCCAAA CAGGTAACCC AACTGACCAC AAGAAAAGCA GCCTAGATCC
25851  TGAGCATTCA GCTCCTGTCT TCACACAACA GACACCACCT CAGTCCCATC
25901  AAAGCCTGTG AAGTTTCCCT ACATCCACCA TTGAGACATA TTCCAGAGCA
25951  GCCTCTCAAA ATTGCCTTAA CAGGATGGGA CACGATATGG TGGAGCTCCT
26001  GGCTCAGGAC AGCTGCCTCA CCCCTTCCTA CTGAGAAGTC TGTATCTGCT
26051  GGTTAGAGCT ATCAAACTGT AGAAGGCTTA GTGCCTGTCC CAGCAAGTGT
26101  CCCCTCAAAA GCCTTCTTGT TTTCTTTCCT TCTGAGAAAA GCATACAAGA
26151  ATGAGACCTT CTATGTTAGA GAGAACTCAG CCTCCACTCT CAATTGACTT
26201  GGTTGACTGA TGAATTGATG CCCTGAGGAG GGGATAGATT CAGGGAAGAG
26251  ACTGTGCTGA ATGAGTCTGT GTTTTCCTAG CTTTGCTGTC TGTGCAAATA
26301  GTGGAACCCA GAAAAATATC GGGTGGTAGA CACACAGACA CTCTAATTGT
26351  CTGAATTTAA ATATTATTTA AATGGAACTT ATAGTATCAT TATATATTGA
26401  TACCATAATA TCACATAAAA TTTGTTTGAT ATATAAACAA ATATTGATAT
26451  TTTATCATAA TATCATAAAG CAGTTGTGCA CACAATAGCA GATAATATTC
26501  TCCGGTTCTA TAAAGTTTAT ATGTTAATGT TCTTACAAAA TTTAACTCAG
26551  TTATTTTATA AAATTACCTA ACAAAATTTT GTTACTGTGC TATCATAATA
26601  ATACATAAAA CTATGAAATC ATAATATATT GTAATATAAT ACATATGAAT
26651  TATGATGTCA TAATGTATTA TATCACAATA CATATGAATT ATGCCATAAT
26701  ATATTATGTC ACAATACATA CAAATTATGT CATAATATAG TGTGACATCA
26751  TGATGCATGT GAATTATGAT GTCATAACAT ACTGTGATGC CACAATACAT
26801  ACAAATTATA TCATAATATA ATGTGATGTC ATAATATATT CATTTATTAT
26851  ATTTATGATT TTATAATAAC ATGAAATCTT GTCAGTTAAT TTTATAAGAA
26901  AATTGAGTTA AATTTTGTAA CAACATTAAC ATATGTAGAA GCTTACAATC
26951  ATGGTGAAAG ATGAAAGATG AGCAGGCATC TCACATGGTA GGAGTGGGAA
27001  CAGGAAAGTT GGGATAAGGA TACGCCTCAT TTTTAAACCA CCGGATCTCA
27051  TGAATACTCA CCATGACAAG AACAGCACAG AGCCATGAGG AATCCATCCC
27101  CATGATTCAA ACACCTCCCA CCAGGCCCCA CTTGTAACAT TAGGGATTAA
27151  AATACAATAT GAGATTTGGA ACAAATATCT AAACTATATC GTATGACCAT
27201  TGGAAAAACA GATGAAACAT TCATGGTTTC TACTGTCCAG ATACTTTCAT
27251  TCCAGAGCAA ATGGCTAAAT GATTGCATTC AACATTCTGA GGTCAGAAGA
27301  GAGAAGGGAG GTGTACAGGG GACTTTGGCT GCATTTGTTC CACTTCCCAT
27351  ATGCTGTTGT TGTGAGTTCT AACATTATCA CCAGAAAGGC TATTCATGGA
27401  CAGAAGAATT ATTGCTATTG TTGTGATTAT TTCTATTTCT TTTACATTAG
27451  TAAAAATAAT TTTTTTAGCT TCTCATATAA TTTTCCTAAA AAAGCCCTAA
27501  GAGTTTTCGT TAAATTCCTT GTTATTGTGT GTCATAAAAA TTGACAGGGA
27551  AATGGCTAAA ATAGATTAAA ATTACACAAA CTCTAGGAGT CAATTCTATC
27601  AGGCAGGCTT AGGAAAGACA GAACTGGAAA TACTCCACCA GCAGAACACG
27651  AGATGCATAG GGCCCACTTT CTGTCCCTGC TGTCCCCAGG TCCACCCTCT
27701  TCTAAGGCCT CCTCCAGGTC TGGCTTCACC CTAGAATCTC CTCTCACAGA
27751  ACTAATTAAA GGAGATCAGA AATTTGAGTG GTGACTCCTG CTGCCTCTCC
27801  TGCGCTGGTG CCCCCAATTT CCTGCAAATA AAAAGCAGAT AAATGGGAGC
27851  AAATAGTTAT CTATTTGTGG GCCACAATTT CTTTTTCATT GAAGCCATAG
27901  AGACATCCCA TCTAGCAACC TGTTTTTTAT TTTTGCATAT CCAGTAGTTG
27951  CTCTACAAGG CACAAGAAAG TCAATATAAA TACCAAAAAA TCCCTCTGGA
28001  ACAGTTCATC CTTTCTTTCT GTATCCCTTT CATCTGTCTA TACAGCTTTT
28051  ATTCCATACA TTTTTCTTTT TAAAGTGAAT AAATTTAAAA AGTGAAAGAA
28101  AAAATAGAAA TGCTGGGCCC TTCTTCTAAA TCCTAGAAAT TATGGAACAC
28151  AGCACCCACT TTCCAGGGTG TGATGAGGAT TCACTCACAT CATGTAAAGT
28201  TTCCAGCACA GTGGTCTGTA ACAGACTTCT GAACACATAG TACATGCTTA
28251  ATAAGCATTG TATTAACTCA TGTGTACATG TTTTTTTTTT TTAATCCAGA
28301  CTTACTCAAA TGTTGATGCC TTCTCTAGCC TCTGTAATCT TTAAAGAACT
```

```
28351  GGCAAAAAAT GACATGTTTG TAAATGGTGA TTGGTGGTGT CCACGTTGAG
28401  CCAAAACTCT CTGTGCTGTG GTAAGAGTGT TGGGTATCAG GAATTCTGTG
28451  TGCTGTGCCT ACTTTCTCTA GATGATTGCT ACTACCACGG ATGTAGTGGG
28501  AGAAACATCA GCATTAAGAG GAGAATTTTT AAGAGAAGCT ATTTCACGGA
28551  CCCCTTTCCA AAACTGCAAA ATCACATCAA CTAAGAAAAA GGCCTGGAAT
28601  CAGAAATGAT TTATATGTGC ATTGAAAATT GAGGGGCAAT ACTTCTCTAG
28651  GTGGCTCTCA GAATAGGCTT TCAACTATAA TCACATGACC CCTTTAAAGA
28701  CCTACCCATA ACCTGTACCC TCCACCACAA GATCCTGTCT GTTGATCTTG
28751  GGTGGGAGCA TCCATATGGT TTTAATTAGG AAGTCAAATT GTCCCTCTTT
28801  GATGATTACA TAATATTATA TCTAGAAAAA TCTAAAGACC ACCAAAAACC
28851  TTTTAGATTG GATAAATGAA ATTTAATAAT GTTTCAGGAT TATTAAAAAT
28901  CAATGTAGAA AAATTAGTAG CATTTTTATA CACTAATAAT GATCAAGCTG
28951  AGAACCAAAT TAAAAAGTCA ATTCCTTTTA CAATAGCTAC AAAAGTACCT
29001  GGAAATACAA TTAATCAAAC AGGTGAAAGA TATCTACAAG GAAAACTACA
29051  AAACATTGAT GAAAAAAATT GTACATAATA CAAACAAATG AGAAAAACAT
29101  CCCTGCTAAT GGATTGGAAG AATTATCATT AAAATGACCA TATTGCCCTC
29151  CAAAAATCTA CATATTAAAT GCAATTCCTA CCAAAATGCC AATGTTATTT
29201  TTCATAGAAT TAGCAAAAGA ATTAAATTTA TTTGTAACCA TAGCAAAGTC
29251  TGAATAGCCA AAGCACATTT AAGCAAAGAG AACAAAGTTG GAGATATTAC
29301  ACTACCTGAC TTAAAATTAT ACTAGAAAGC TTGAATAACC AACACAACAT
29351  GGTACTGATA CAAATAGACA CCCAGATCAA TGTAAGAGAA TAGAGAACCT
29401  AGAAATAAAG CCACGTACTG ATCATTTACA AAATCAATAA AAGCATACAT
29451  GGAAAAATGA CATTCTATTC AACATGTTGT GCTTGAAAAA TTATATTACC
29501  ACATGCAGAA GTATGAAATG GAACCCGTGA CTCTCACCAT GTAAAAAAAT
29551  CAACTCAACA TGGATTAAAA GACTAAAATG TTAGACCTGA AATGATAAAA
29601  ATTCTAGAAG AAACCCAATG ATAAATGCTT CTGGACATTG GCCTGGGCAA
29651  ATAATTCATG ACCAAGATCT CAAAAGCAGA TGTAGCAATA ACAAAAATAG
29701  ACAAATGGAA CTTAATTAAA CTAAAAAGTT CCTGAAAAGG AGCTTTTAAT
29751  TAGCAGGTGA GCAGACAACC CATGAAATGA GAAAAATGTT TGTGAACTAT
29801  ACATGTGACA AAGAACTAAT GTCTACAATA TACAAGGAAA TCAAACAACA
29851  AGAATAAAAC AAGTAACCTC ACTATAAAGC AGGCAAAGAA TGAGAACAGA
29901  TATTTTTCAA ATAGAAGACA ATGATTGCCA AGAAGCATGT AAAAAATGTT
29951  AAACATTGCC AATGATCAGA GAAATTCCAA TTAAAAATCG CAATGAAGTA
30001  CCATTTTATA CCATTCATAA TGGCTAATTA TTAAAAAGCA GAAAAATGAC
30051  AGATACTGGC AAGGATACAG AGAAAAGAGA ACACTTATTC ATTGTTCGAG
30101  GGAGTGTAAA TTTCTACAAC CTCTATGGCA AACAGTATAA AGATTTTTCA
30151  GAAAAACTAA AAATGGAATT TCCATTTGAT CCTGCACTTC CTCTACTGGG
30201  TATCTACCCG AAGGAAAATA ATTCATTACA TAAAGAAGAT ACCCACACTC
30251  ATATGTTTAT TGCAGCACTA TTCACAATAG CAACGATATG GAGTCAGTTT
30301  AAATTTATCA GTCAATGATT GGATAAAGAA AATGTACTAT ACATTTATTC
30351  CATGGAAAAC TACTCAGCCA TAAAGAATAA AATCATGTCT TTTGCAGCAA
30401  CATGAATGCA ACTGGAGGCC ATTATTGTAA GTGAAATAAT TCAGAAACAG
30451  AACATAAAGC ACTACATTTT CTTACTTACA AGTGGGAGCT CAATAATGCA
30501  TACACTTGGA CATAGAGATT GGAAAAATAG ACACTGGAGA CTCACAAATA
30551  TGGGAGGTTG GTAGAGGGGT TAGGAATGAG AAAATACCTA ACTGGGAAAA
30601  TGAGCACCGT TCAGATGATT GTTGCACAGA AGCCCACACT TCATCCCTAT
30651  GCAACATGTC CCTGTAATGA AGCTACATTT ATACCCTCTA GTGTATTAAT
30701  AAAGAGAAAA AAACTGACTT TTTTTCAGTT TTGACAAGAG GGCAACTGAC
30751  AAGAGGGCAG ACTGAATAGA CTTTATAATT TTCATAAATA TTTATATTAG
30801  AAAAATAACT TTAATAAAAA TAAAAAATAA TATTGTATTT TTAAGAATGG
30851  TATAAAAAGA TAATTTGATG AATTGGAATA GTTAGGACTT AGCACATACA
30901  ATATGTTAAG CAAGATTCTA AGCCATTAGA CATTTGTAGA CAGAATATCT
30951  AACAGTATAA AATAAATAAC ACAAATATTC ATTGGCAATG ACAAATTGAC
31001  ATATTTTCAA TCCTATTAGA TGATATTAAA ACCATTACAA AATTTACTGT
31051  TTTGTTTCAT AATAAAAATC GTAATGTTAA ATAATTTCAT TTAAAAGTTT
31101  GACTAATTAG GCATATAGAG AAATGGGCAG TATGTTGACC AGTAAACAGG
31151  ATACAAATTA TTTTTCGAAA AGCAAACAAC ATTTTTATAT TTTATTTATT
31201  TATTTATTTT TCTGGTCGGA GGAGCTGCTT TATTGTCTGA GGACATAGTA
31251  CGGTCCTCTC CCTGGGAGGT GGGGTCTTCC ACCGGTCACC CGAGGCAGTG
31301  TTCCAGGAGG CTCCATGCAA CTCACTGCTG GGCTCAGCTG GGGGCTGGGC
31351  CTTGGAGAAG GCGAACTGTG CAGGGAAGCA GTAGCTGTGG GTCCTCACCG
31401  CCCGCTCTGC TTTGCTGCAC TGGGTCCCTG GTGCTCCTCG AAGTCCCACT
31451  GAGCCTGCGT CTCTATAGGA CGGTGACCAT CCAGCCCAGA ATCTTCTAGT
```

```
31501  CAGAGCACAG TTTGACCAGG CCAGGCATTT CCGCTTCCTC TCCTTGGGCT
31551  GGACTTTGCA CTTGGGTTTC TTCCAGTCCT TCTTCTGCCG CCTGTCTGCC
31601  AGAGCTTAAA TTCCAGCCTC ACAAATGTTC CAGCTAGAAA AGGCGTGTCC
31651  GCGGCGCCTC GCACACTGGT CTCTTGGAAG GCACGGGCGG GTGGATTCCT
31701  CCAGGGCCAC CTGCAGGTCC GGGTGCTGGG CCCTGTGAGC TCGACCCCAC
31751  CCCTGCCCCC CGAGCCCATC CACTGAGCCA ACGGTATCCT CAGCCTTCAC
31801  TTGCTTCCAC CATCACACCG CCCTACAAAG CGGGTGTGCG CTCCTCAGCT
31851  CTCCCTGCAT GCCAGGAGCC GCCTCCTCCC CTACCCTGGC CCTGGGGTCC
31901  ATGCCCACAG AACGCTGGGC AGAGGTGAAG GAACTGGGAA ATACCCCTTT
31951  CTCAAATGAC TTTGGGGTCC GCCTGGTCTT CTCTACTCCC TCCCACCCTA
32001  CCTGCACTGT TCCCTGGGGC CTGCAGTTTT AGCAAAGTTC CCTGCCCCCC
32051  ACCCGTGTCA GGAAGCAGTC CTGATGCCCA CTCCCACCCT TTCCCCTCTT
32101  CTGGCCCATT CTCTCTCCCC ACTGGGTCAC TGACAGGAAT CCCTCTCCTC
32151  CCTGCTGTCC CTGGAGCCTG TCTGCTTTCC GCGCTCAGCC TCCTCCCTGA
32201  CCGCTTCTCC CTCCTTATCC TGAATCTCCC GACTCCCGTG GGGTGGCCCC
32251  CCCACCCATC CCGGGGCCCA GGCAAAACCG ACAAAATTAT TTCAAATGGG
32301  AAGATCTGAA TTCCACTGAA GTCATGACAA CAGAAACCCT TGTGGTCATA
32351  TTTTAAAGAA TGATAGTAGC AACTATCCAT GTAAAGTATA ACTAGCCTTA
32401  AAAATACCCA CCACTTTGCC AAAAACAACA ACAAAACCAT GTTATAGCTG
32451  AAAAACAAAT TCAATAAACT TACAGGATAC AATATTAACA TGAAAATCAG
32501  TTGCATTTTT GTACAGTAAC AACAAAGTAT CTGAGAAAGG AATAAAGAAA
32551  ACAATTCCAT TTACAATATT ATCAAATAGA ATTAAATACT TAAGTCATTA
32601  AATAGAATGA GTTTAACGAA GAATAGTAAA GATCTGCATA CTGAAAACTA
32651  TAAAATGATG ATAAAAAATT GAAGAATACA AAATGGAAAA TATATCCTGT
32701  GTTCATCGAT TCTAAAAATT AACATTGTTA AAATATCTAT ACTACATAAA
32751  GTTATGCTAC AGTTAAATAA ATTTCTATCA AAATTTTAAT GCCATTAAAA
32801  TAAATGTAAA ACAAACATTT GTAAAATTAG TATGGAGCCA CAAAAGACCC
32851  CAAATAGCCA AATATTGAGA AGAATAAAAA GGCTGAAAGC CTCAAACTTC
32901  TTGATTTCAA ACAATATTAC AAAGCTATAC TCATCAAGAA AGTATAGTAC
32951  CTACATAGAA ACAAATGGAA TAGAATAGAG GACCCAGAAA TAAATTCACA
33001  CATATACAGT AAACTGATCC CACCAAATTA GGAAAAGATA GACACATCAA
33051  GAAATGGTGT TAGAAAAAAC TAGCTATGCA CACACAAAAA AGTAAAGCCT
33101  TCTCTTATCA CAAAAATGAG TTTAAAATAA AGACATAAAC ATTATAACTG
33151  AAATAATGAA TCCTCTAAAA AAAAAAACAG GGAGAAAGCT CCTTGACACT
33201  GGGGGGGGCA NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
33251  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
33301  NNNNNNNNNN CTAATAATAA CCTGTACTTG AAATTTGTTT ACAAAAGTAG
33351  ATCTCAGGTG CCCTCACCAC ATACACAGAA ACGTATAAGG TAACTATGTG
33401  TGGGGATAGA TATGTTAACC AGCTTGAGGT AATTTAAAGA TATATACACA
33451  TCTAAAAGCA CTCTATTGTA CACCCTAAAC ATACATACTT TTCAATATGT
33501  AAATCATAGC TCAATAAATG TGGCGAAAAA ATTAAGAGTA ATCAGTAGGT
33551  CATATTTAGC CACAGGTACA CTTTATTTAA AACTCTCTTG TGAACTGTTT
33601  CTGGCTGAAC CTGGGAACGC CCCCAGTACT TCTGACCCCT TGACTTCTCC
33651  AACACTACTT CTATTCTTCC TGTAGCTGGG GAAGATGTGA TAGTTCCCAA
33701  GAGTTTTATG GTTTCCAGGC CATGCAGATC TGGGAGGGCC ATCAGCCGTC
33751  CCCTTGCTCA TTGCGCACAG CTCCAGGCAA CCCATGTCAG GGCATCGCAT
33801  CCCGGCGAGG GCCACAGCAG ATAGCAGGGG CGAGGTAGCT CTTGGCTGCC
33851  TGCAATCCAA TGGCCTCTGG CTCCACAGGC GTCCTTCAAG GCCCTACCAC
33901  AGCCCCTCCT CTCATGACCA GTGAAGGCAA TGTAACTGCA GAAGACACTG
33951  AGGAGGCAAG TAAGAAAAAA TCTTGGCCCC CGTGTATTAA TTATCATTAG
34001  ATCACCCTTG TTGTATACAG GATCACTCTT TATCATTACT GCTGTAAAAC
34051  CTGCTACAAT TGGAAGAGAA AATGATCTGG AGCATACTCT TTGAATTCCT
34101  GAAGTTGCTG TCAGTAAGTT TCAGCAAAAA TTTATTCTGA CCATGATTTA
34151  CAAAGTTATG TCCTTGTATA TCAAGGCAGT CAAAATGAAA CAATTGTTAA
34201  TGGAAAACGG ATTGTTCGGC TGACAACTAA ATGTGAGCCT TATGTACCTA
34251  AGCATAGACA TGAAATGAAA ATTGGAGAGA CTAGGTTACC CTTTCACATT
34301  CACCCTGGCA GTGATACTTG TGATGGCTGT GAACCACAGC AGGTTGGAAC
34351  TCACTTTCAC CTTGATAAGA GAGAGAATTG TTTATTGAAT TAATTGTTTA
34401  CAGCTAAGTA GGGGAAAAAA GAGTTGGAAA TAAGAAAATA ATTTTTAAAA
34451  TGTGAGTAAA GTATGTTTTA CATTATACAG ACTATGAAGA TGAATAGACA
34501  TTGAAGAACC TGAAACATAA AGATACAGCT GGAAAAAGTA AGGAGCCTAT
34551  TGGAAGTGAA GAACACTTCC AAAGAGTTGA TGCACCTGCA TCTGTTCATT
34601  CTGAAATTAC TGACAGCAAC AAAAGTCAGA AGATGTTTGA GGAGGTGGGT
```

```
34651  TGGAAAAAAG AAGAGGGCCT GGGGAAGGAT GGTGGAGGAA TGAAAACTCC
34701  AATTTTGCTT TAGCTTTAGC AGACACACGT AGGCTTGAGG ACAGGCAATT
34751  CCTCCTCAAT TGAAGATGTT CACCTTCTCC AAAACAAAAA CAAACTGGGA
34801  CAAAGCATGA GATAGGTTTT CTGAAAATTT CCCAGAAACT AAACCTTGAA
34851  AAGATGACCT AGGGACTAGG CCTTGGGTAA AAAGGGGCTG TCGAGTGAGG
34901  GTTAGTCATA GAAGAAAACT CAAGTTTTTT AAAAAATAGA GTTTGGAAAC
34951  TCTTATTTAT TTTATTTTAT TTTTTGCAGA ACTTTTCTCC CAAAAAGAGT
35001  CTGTGGCACA GTTTACCCCT TCCTGATTCA GAAATGTGTA ATAAAGTTTG
35051  GTTTGCAACT TTTCAATGCC ATTTTTTTAA ACTAATAAAT AGTGATTTAA
35101  TCAAGTTATG CAGTAAGTGG ACTAAAGTTT ACAGGGCACA CATGAAGTGT
35151  GTCACACTTC ATTATTTTAT CGTGTCGTCT ATGACATCCC TGTGAGCACA
35201  AAGCCCTCTA TGCACAATTC ATATTACCAC TACTGACGTC AATATACATC
35251  TTGTCTCTGT CTCCTCCTTT CCCAGCANGA CCCCCAGCAG NAGAAATATT
35301  CAAAGTGTTA AAATAAATCT GCTGTATGCA TCCTNNNNNN NNNNNNNNNN
35351  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
35401  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNCCATTT ACAATATTAT
35451  CAAATAGAAT TAAATACTTA AGTCATTAAA TAGAATGAGT TTAACGAAGA
35501  ATAGTAAAGA TCTGCATACT GAAAACTATA AAATGATGAT AAAAAATTGA
35551  AGAATACAAA ATGGAAAATA TATCCTGTGT TCATCGATTC TAAAAATTAA
35601  CATTGTTAAA ATATCTATAC TACATAAAGT TATCTACAGT TAAATAAATT
35651  TCTATCAAAA TTTTAATGCC ATTAAAATAA ATGTAAAACA AACATTTGTA
35701  AAATTAGTAT GGAGCCACAA AAGACCCCAA ATAGCCAAAT ATTGAGAAGA
35751  ATAAAAAGGC TGAAAGCCTC AAACTTCTTG ATTTCAAACA ATATTACAAA
35801  GCTATACTCA TCAAGAAAGT ATAGTACCTA CATAGAAACA AATGGAATAG
35851  AATAGAGGAC CCAGAAATAA ATTCACACAT ATACAGTAAA CTGATCCCAC
35901  CAAATTAGGA AAAGATAGAC ACATCAAGAA ATGGTGTTAG AAAAAACTAG
35951  CTATGCACAC ACAAAAAAGT AAAGCCTTCT CTTATCACAA AAATGAGTTT
36001  AAAATAAAGA CATAAACATT AGAACTGAAA TAATGAATCC TCTAAAAAAA
36051  AAAACAGGGA GAAAGCTCCC TTGACACTGG TGGTGGCAAC GATGTTTTGG
36101  CTTTGACAGA AAGAACACAA GCAACTAATG CAAAAATAAA AAAGTGGAGC
36151  TATATCAAAG TAAAAACTTT CTCCACAATA AAGAAAACAA TCAACAAAAT
36201  ATAAAGACAA TATATGGGAT GGAAGAAAAT ATTTGTAAAC CATATGTAGG
36251  ATAAGATGTT GCTATCCAAA ATATACATAA CACTAGTCAA TGTGGAAAAA
36301  AAAACTCACA GCAGAACAAA ACAAAAACCA ATTTCCTGAT AAACTGGGCA
36351  AAATATCTGA ATAGATGTTT TCCCAAAGAC ATACAAGTGG CCAGCAGGTA
36401  TATAAAAAGA TTCTCAACAT TGCTAATTAT CAAAGTAATG AAAATCAAAA
36451  TCACAATGAG ATATCACCTC ATCGTGGTGT TAGCACAGCT ATTATCAAAA
36501  ATTCAAAATA CAAAAAGTGT TAGGGTGCAG AGAAAGGAGA ATATTTGTCC
36551  ACCATTGCTG AACATGCTCA CTGGTGCAGC CATTATAAAC AAAAAAACAA
36601  CACAAAAAAA CAAACAAGAA AACCAGTAGG GAGGTTCCTT AAAATTTTTA
36651  AACTAGAAGT ATCTTTAATC CCAATTTGGA GTATACAGCC AATAGACATA
36701  AAATTAGTAT CACCAAGCGT ACCTGCACTC CTCTGATACA AATAAATAGG
36751  TAAACTGTGA GAAAGAGATA ATTTCAGCTT TAATAAAGAA TGAAATTGTT
36801  TCATTTACAA CAATGTTGAT GAACCTTGAA GACATTATGC TAAGTGAAAT
36851  AAGCGAAATA CAGAAAGACA AATACTGCTT GATCTCATTT TAATGTGAAA
36901  TCTTAAAGAA AGAAAGAAAA AGAAAGAAAG AAAGAAAAAA TGAAACAGAG
36951  ACTAGAATGG TAGTTACCAT GGGCTAAGAA TTGGGGAAAA GTGGGGAGAT
37001  ACCCATCGAA GGGTGCATAC CTTCAGTTAC ATAATAAAAA ACTTCTGGGG
37051  ACCTTATGTG CAGAATGGTG ACTATAGCTA ATAATAACCT GTACTTGAAA
37101  TTTGTTTACA AAAGTAGATC TCAGGTGCCC TCACCACATA CACAGAAACG
37151  TATAAGGTAA CTATGTGTGG GGATAGATAT GTTAACCAGC TTGAGGTAAT
37201  TTAAAGATAT ATACACATCT AAAAGCACTC TATTGTACAC CCTAAACATA
37251  CATACTTTTC AATATGTAAA TCATAGCTCA ATAAATGTGG CGAAAAAATT
37301  AAGAGTAATC AGTAGGTCAT ATTTAGCCAC AGGTACACTT TATTTAAAAC
37351  TCTCTTGTGA ACTGTTTCTG GCTGAACCTG GGAACGCCCC CAGTACTTCT
37401  GACCCCTTGA CTTCTCCAAC ACTACTTCTA TTCTTCCTGT AGCTGGGGAA
37451  GATGTGATAG TTCNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
37501  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
37551  NNNNNNNNNN NNNAAGGATG GTGGAGGAAT GAAAACTCCA ATTTTGCTTT
37601  AGCTTTAGCA GACACACGTA GGCTTGAGGA CAGGCAATTC CTCCTCAATT
37651  GAAGATGTTC ACCTTCTCCA AAACAAAAAC AAACTGGGAC AAAGCATGAG
37701  ATAGGTTTTC TGAAAATTTC CCAGAAACTA AACCTTGAAA AGATGACCTA
37751  GGGACTAGGC CTTGGGTAAA AAGGGGCTGT CGAGTGAGGG TTAGTCATAG
```

```
37801  AAGAAAACTC AAGTTTTTTA AAAAATAGAG TTTGGAAACT CTTATTTATT
37851  TTATTTTATT TTTTGCAGAA CTTTTCTCCC AAAAAGAGTC TGTGGCACAG
37901  TTTACCCCTT CCTGATTCAG AAATGTGTAA TAAAGTTTGG TTTGCAACTT
37951  TTCAATGCCA TTTTTTTAAA CTAATAAATA GTGATTGAAT CAAGTTATGC
38001  AGTAAGTGGA CTAAAGTTTA CAGGGCACAG ATGAGTTTGT CAAACTTCAT
38051  TATTTTATCG TGTCATTTAT GACATCCATG TAAGCAAAAA GCCATATAAG
38101  CAAAATTCAT ATAACCACTA ATGACTTAAA TATACATTTG TCTTTGTCTC
38151  CATATATTCA CAGTAAGACC CACAGCAAAA GAAATATCAA AAGTTTATAA
38201  AAATAAATCT GGCTATATGC ATTCTTGTTT ATGCCCTTTA GAACCTAGAT
38251  AAAAGGACCT TTATAATAAA GGTCTAAATA ATACCATTTA AAGCCTAAAT
38301  AATACTATTT AGGCTAAATA AATAAATAAA GGTCCAAATA ATACTATTTA
38351  AAGCCTAAAT AATACTATCG AAGAACTAAC GAACAGGTGA CATACTATAG
38401  AAAAGTAGTC TTTTACTGTT TTCTTCTGTA AAGAATCTGT TGTTTTGTGC
38451  TATATATTCA GCATTTATAT TTCATTTGTT TCATAGCTAA TGAAATATTT
38501  AGATATGAAC AACTGAGTAC AGTACTGAAA TAGTGCACTG GCATTTGTAA
38551  TTTTTATAAA TATTATTGCA GGCAGTGGAG TTGTGCCAGA GAAATCTGAT
38601  TTCTAGTACA AAAGGAATAC TTAGCTAGGG CCTGAAGTTT AAGATATTTA
38651  TTGAACATGT CCTCAATTGC AATATAAACA TTATAACATT TTTTAAAAAT
38701  TCTTTTTAAT ACATTCTGAA TTAAACAAAA ATTTCAAATC ATACCTTTTA
38751  TGAAGTTAGG GAAGTGCTAA TGAGGTAGAA TGGCAGTTGA AGCCAAAACA
38801  TCTGAATTTA TGTAAATAAT TTTACCCACA TTAGTTTTTT GTTAAAGGAA
38851  ACAATAATTC TCACCATATA CTATTTACTG CAGTGAAGTA TTTTCAGAAA
38901  CGTGTGCATA ATACATAATT AATTTTCTAA TGGTAATACA TTCTACAAAT
38951  TATTACAATT GTGGTTTCTT GAAGAGAGCA GTATTTCAAA TGAAGAAACT
39001  TGGAATTTCT CATGAGGGGA TGATCATGAG GATGATCATG AGGGGATGTC
39051  AAATCCATGT GGGTTGTTCA GTTTCTAACA AGACAAATGG AATAAACATA
39101  GATGAACTGA CTGCAGGCAT CCCCAGTGTG CTGTCTTCCA TATGTGAGAA
39151  CAAAATTATA AATTATGTTA CACAAAAGGA GATAAGTTAC TAGTGTGAAG
39201  CGTTAGATAA ATTATATGGC ACATAAAACT GAGATATGAG GTTTGTATTT
39251  ACATGATTGA CAATATAATA AAAAAAGAAA CAAAAATAAT GTGGAAAAGA
39301  ACTACGAGAT TGTAAAATCA TATTTGGAAA AGAGGCAAAT GGAAAGTAAA
39351  CTATTGAAAA TGTAGTAAAA CTACTGTACA GCATACAAAT TACACATTAA
39401  AATAGGCTGA GCTGATGAGT AGGACTAATA AACTGAAACG CTGAGCACAA
39451  TTAATGTAGT CATATACTTT TTAGAAGGCA AATTAATACA AAATACAAAT
39501  ATATTTATTT ATATATGAAG ATTAGCTGGG AAGAAATGAA AAATAATTTG
39551  TTTTACTAGA CTATATAAAC AGGAAGGTAA TATTCAAAAA ATCAGTGACT
39601  CATAAGTTTC AGAAATTGGA AACACATGAA TCCTATCAAA GAAATAATGG
39651  TTTTGGTTAT TGTGAATACT GCTTCAATAA ACATGGGAGT GCAATTGTCT
39701  CTTTGACATA CTGATTTTAT TTCCTTTTGA TATATACCCA GTGGTGAGAT
39751  TGCTGGGTCG TATATATGGT AGTTCTATTT ATAATGTCTT GAAAAATTTC
39801  CATGCTGCTT TTCCTACCTG GCCGGGGCTC CGACAGCTGG GCATCCGGCC
39851  GCAGTCCCTC TCTCAGGATC CCTCCACCCT CCGCCTCCCA ACAGTTCGGG
39901  CTTTTGTGTA CGCTGTGGCT GCTGCTTCTG CTGCCGAAGC TTGGCATTGG
39951  AGACACCTCG TCCTCCTCTC AGGACAGATC CATGAACCCA TCGGCAGCGG
40001  CGGTGAGCGA TGCCTTCCCT CTGCCACAAG GCGCCGCCTG CAGAGCCTGC
40051  CGCGTCCGCC ACGCCCAGAG CGTACCCGCA GCTCAGCGCC GAGTTACTCC
40101  TGCTGGCGGT GGCCAGGGAG TGACTGGAGT CGCCGGACAC CCCTGGGGAG
40151  CAGGCGAAGG AGGAGCTGCA GCCACCACGT CCTCTCTTTC CCCAGGGATG
40201  TGCAGAATTA CCATGAAATT ATGACTCCTC ATCCTAAGAA TTACCAATGG
40251  GAAAATTGGA GTCTAGAAAA TGTTGCCATC ATTTTAGCCC ACCGGTTCCC
40301  CAATAGCTGT ATTTAGGTGA TAAAGTGCTC TCGAATGCAT TGCACAGACT
40351  CAGTTGCCAT GACGATTTTG TGAAAAGTAA CATGTTTGGT TCCCCAGAAC
40401  ACAATACTGA CTCTGGAGCT TTTAAGCACC TTTGTATTAT TAGTTAATGC
40451  TTTTAAGTCA TAATAGTTTA TCAAAGAAAA ATTTGAACGG TTGGAATAAG
40501  GACTCCACGC ATGTAATTGC AGTCCAGTTT CCTGATATTA CAAATCGTTT
40551  CCAGGGAGAA AAAGAGAGGA CCTGTGAAAA ACCTGATGAG TTGGCTATGA
40601  GTTTTTATCC ACCATCACTA AATGATGCAT CTTTTAATTT GACTGGATTC
40651  AATAAAGATT GTGTTGTTTT GAATCAGTTT CTTTTTGAAT TGAAAGAAGC
40701  CAAGAAACGC AAAGACATAG ATACTTCCAT TAAAAGCATA AGAACAATAT
40751  ATTGGCTGGA GGGTGGTCAT TCCGTAGGAA GTAATACCTG GGTTACTTAT
40801  CCACAAGTCT TGAAAGAATT TGTACAGTGA GAGATTATTG TTCACACCCA
40851  TGTAACACTT CACCAAGTAC ATGATCCAAA GAGATCTTGG ATTGGAAAAG
40901  AGCAAAATAA ATTTGCTTAG ATACTTGGGG ATATTGGTAT GCAGGTGACT
```

```
40951  AGCTGAATTC ATTTCATGAA GGAAGCTCCC TGTATAGAGA ATCCCTTTAG
41001  AGTTCATGAA GTAGTTTGAG GCTACAAATA TATTGATGTA CTTGTTCAGT
41051  GGAAGAGCAT AAGCACTTTG AGTGTTATGA ATTCAGATAA TGGAATGTAA
41101  TTCATAGGTG CATTGTCAGT ATGGGGGAAA CACACGTTCC TGAAATATGA
41151  GTGAAATATG CAATAGTATT TCTTCCTTGG GAATGTGAGC AGTTTTTAAT
41201  TTGTGTTGAG TTAGAATTAG TTAATTTAAA ATCTAACAAG GTGGTTTGTG
41251  ATAATACTGA GGAGATATAA GACCCTTAAA AGGAAAGTTA CAACATAGTA
41301  CTTCTAGAAT ATAACTAAAA TTGTTTCTGT TGGAAATAGT GATTCTCTGA
41351  GTAATGTTAC TAATCGTGGT GATATTTTAA CAGTAATTAG CTATTTTGGC
41401  ACTTAAAACT TGAATGGAAA CAGTTTATTT CTCTTCAAAC AAAAGCAAAG
41451  GCACAATGTT GTTTCCTATC ATTTTGGAAT AACTGTACCC TGCCTCTTGT
41501  GTTTTGTAAA CTCATTCACT CATTCTTTAA TGTGCCACCA AGTACTTTTT
41551  TCTTGAGAGT CAAAATATAT TTGTTTCACA ATGTCCAAAA ATGTGCAAGA
41601  ATGTAAAGCT GGTTTTTAAA AACATAGCCA TGTGATGGCA TGTGCCGTTA
41651  GTCCCAGCTA CTCAGGAGGC TAAGGCAGAA GGATCCTTTG AGTCCAGGCT
41701  ATAACGCACC ATGATTGTGT TTGTGACTAG CTACTGCACT CCAGCCTAAG
41751  CAACATAGTG AGACCTCATC TCAAAAAAAA AAAAAAAGAA AAAGAAAAGA
41801  AGAAAAAAAA GAAAAACAGA ACAAATTAGA CTTAAGTACC ATCACTTAAT
41851  TTTTAGTTGA CAGTCTTTAG TTGATTGTTT TGGATAAGAC ATTCTGGGGC
41901  TTCTTGAATC TTGGCCAAAA ACCAGTTGTT TTTGAAAACT GTTTTAAATT
41951  AAGCATATTT ATGTATTTTG GATAAAAATC AACTACAAAG AAAATTTTAT
42001  TTTTTTCATT ATATTAGTCT TTTTGAAAGA GAACAACTTA GGGAAGATAA
42051  ATATATAATG CTCTATTTGT CAATGCTGTA TTAAAAAGGA AACAGATTTC
42101  ATAAATCTAA ATCAATGTTT CTCCACAATC ATGACTTTGT CTCAAAAAAA
42151  AAGTTATTTT TGGCCAAAAT GCAAAATTAT ATTGCTGTGA CAAAAGTCAC
42201  AAGGAATCGC TTAAACATCA TCCAGCCTGA GGCCAGATCA ACATATGACA
42251  GTCACAATTT CAACTCTGAA CTGCATCCAT GTGTGAGATT TAGAGTCTCA
42301  TTAGTGGACT CTGCCCATGT ATGAGGTTGA CAATCCTAAT TGTTTTCTAC
42351  ATGTGTCTAT GAGTGTCACA AGGTCACTGT TTGCTGGGCC CTGTTATGTA
42401  ACTCTCTCTA AACCCCAAGG GGTTTATAAA ATACATGTGA GTGTCATAAT
42451  CTTCTGTGGC CATTTTAGAA TTAGTAGACC AAGGACCTTA GTTGTTGCCG
42501  TAAGCCTAGC TATTAAAGTC AAAATTACTC CTCCTGGCTG GGTCCATGTA
42551  AGAAAGCTAT CATCATGACT GTGGCCTGGG CCTAGGTATA CGTCACACCC
42601  CACCTGTGAG CAGAAACAGG GAGGAAAACC ACATCATCTG AATGCTGGGA
42651  CAGGGAAATG TCAATATTGC TCAATTGGTA GGACCCACAC AGGAAAGTCA
42701  CATCACCTTT GTGCTAGGCT CAGTGATATG TCACAATGCC CACTGTAGAC
42751  AGAACCTTGT CAAAAGAGTC ATATCACCTA GGTGCTGAGC CCAGCAATAT
42801  GTCACAATCC CCCTTGTTAA CAGGGCCCAG TCATGAGAGG AGAGTCCTAT
42851  CACCTAGATG ATATGCCCAG ATTTCTGTTA CAAACCTTAC TGGAGGCAGG
42901  GCACAGGCAG GAAAGTAGAG TATCATCACT CAGGTGATGG GCTCACAGGC
42951  ACATTACAAT GTCCCTTGTG GGTATGATCC AGACAGGACA GTCACATCAG
43001  CTCAGTGTTT GGGTCAGTTG TATGTAATAA TCCCAACTGT TGGCAGGGCT
43051  CATGCAAGGG AGTAAAGTCA ATCAAGTGGT GAGAAAAAAA TTGTATTTCA
43101  AAGTCACACC TACAGGAAAG TCCAGGGATG AGATTCACAG TCCCACACAT
43151  TTTCTGACTC CTAGGATAAG AGACAACAGC TTCTTTGAAT TTTGTTAAAG
43201  TACACAAATC ACAATCTCAA TGGTGGACAG AATTCATGCA TGAAAGCTCC
43251  AAACACACCT GCAAACACTG TCTAATTAGA GCAGCCACGA TCTCACAGGT
43301  GTGCTGAATT TTGATATATA AGTCACCATT CTACCTGTGA ACTGTATCCA
43351  TGTATGAGAG TCACAATGTG AACTTTCAAC TGATCTGGGT ATGATATTCA
43401  GAACCTCAAC AATGGTCTAT GTCCCTATAG GAAGATGACA TTCCTCTCTG
43451  TATGTCGTGT GTGCTTGGAA GAGTCAAAAT GTCACCTGTG TGCTGGGCCT
43501  TTATTAGTCA CCTTCTGTGC CACTTAATGG CTTCATATGG TCTGCATGAG
43551  AATGAAAACC TGCTCTAAGA TTTCATGATG GCATAAACCC ATGATCCTAC
43601  ATGTTGCCCT AAGTTGAGGC ATGAGAGTCA AAACCTGTCT TATTTGCTGG
43651  CTCCATGTAT GAGAGTCATC AGTGTGCCTG TGAACTGGGT TCAGAAATGA
43701  GCCACCATCC CATTTGTGGA TGGATCCAGA TATGACAGTC ACAATTCCAA
43751  CTGGGAAATG TTTCTGTAAG TGAGATCCAG GGCCTCGTGA GTGGGTTCTG
43801  TTCACATCTG TAATCTTTAT ACAATTAGGA GATGCAGAAC CTTACTGATT
43851  GTTCTATGAG AGTAAAAATA TCTTCTACTG GCTGGGCCCT CATATGAGAG
43901  TCTTCATTAT TCCTGTGACC TGAACCTAGG GATATGTCAC AATCTTACCT
43951  GTGAGTAGAA CCAGGCAGGA GAGTCTCATC AGCTGAATAC TGAGCCAGGG
44001  ATATATTAGT GTTCCCCCTG TGGTTGAGTG CTGGTAGGAT ATTCACATCA
44051  CTATGGTGCT GGGACAAGTG ATGTGTCACA ATGTCCCCTG TGGGCAGAAC
```

```
44101  CCAGGCTAAG CATTACATCA CACGAGTGCT GGGCTCATCA ATATGTCACA
44151  ATACCCTTGG GAATGGGGCC CAGGCAGGAG AGTACAATAA CATCATCTGG
44201  AAATGGGATG AAAGGAATAT CACAAAGCCA CCTGTGAGAA GGGACTAGGC
44251  AGGAGGGCTG CATTACTGGG GTTATGGGTA TAGTATATGT CACGATCCAC
44301  ACTGTGGGTT ATAGAGAGAC AATATATTCA TCTCAGTTTA TTTGCGGGTG
44351  GACAAAAGTA TATGTCAGTC ACACATGTGG GAAGGTCTAG AAATAAAATT
44401  TACATTCCTA CAGATGNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
44451  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
44501  NNNNNNNNNN NNNNNNTAAT TTCTGGCTGG ACAAAGTATA TGTCAGTCAC
44551  ACATGTGGGA AGGTCTAGGA ATAAAAGTTA CATTCCTGCA GTATGTTCTG
44601  GGTCCAGGTA TAAAAGTATG CACCTCATAA ACTTTTACAC ACCTCATTTG
44651  CCCAAGTTTG CAAGTCATGA TATCAACAGT AAATTGGATC CATACATGAA
44701  AGACTCAACC TCAAACAACA CACATTAGAA GAGTTAGAGC CTAACATATT
44751  TGCTGAATCT TAGTCAGAGA CTCATCATCT CACCTGAAGG CTGGACCCAC
44801  ATATAAGAGT AATAATACCA CCTTTGGACT GCCTTTGGGT GTGAAATTCA
44851  GAAATTCAAT GGTGGGCTGT GTCCATGTGG GAGTGTGAAA AACTTTACTG
44901  TCAGGAGTGT ATGCAATGAG AGTCAAACCT TCAACACTGT ACTGGGCCTT
44951  GTTTTATTAT TCTTTGCACC ATCTGTGGGT TTTTATACTA CGTGTGAGAC
45001  TTGCAATCCA CTCTGAAGCA TTTGTCCTTG TATAAACCCA TGATCTTACT
45051  GTTAGCTAAG TGTGCAGACA AGAGTCACAA TTTTGTTTGT GTTATGGACC
45101  CTGTTATGAA ACTCTCTTTA CCACCTGAGA GCAATGTGCA ATATGCTTGA
45151  GTGTTGTAGG GCTCTGTGAC TTTTGTACAA TTAAGAAACC CAGAACACTA
45201  CCTGTTGCCC TAAGCCTAGT TACTAAAGCA AACATCTATT TTATTGGCTG
45251  GGTCCAGGTA TGAGAGTCAT TACTGGACCT GTGAGCTGGG TCCAGAAATG
45301  AGTCACCATA CCACCTGTGG CCAGATTTAC ATATAGCAGT CACAATTCCA
45351  ACTGTGGACT TCATCTGCAT GTGAGATTCA GGAGCTCACC ACTGGGCTCC
45401  CTTCATGCAT GAGGGTGACA ATTCTAAATT TCAGCAGGAT GTGGCTACAA
45451  GAGTCACAAC ATCACCTGTG CTCTGGTTTG TATTATGAAA CTCTCTGTAC
45501  CACCCAAAGG CTTCATAATA TACGTGTGTG TGTCAGAATC TCCTGTAACG
45551  TTTTACAAGT AGGAGACTCA GGACCTTACC CTTTGCCCTA AATCTATCTA
45601  TAAGAGTCAA AATATCTTCT ATCAGATGGA TCCACATCTG AAAGGCATCA
45651  TGATTCCTGT GTGCTGGGCC TAGTAAAGGT CACAATCTTA GCTGTGGGTG
45701  AACAGAAGGC AAGAGAGTCA CATCACCTGG GTGACGGGCA AGAGATATAT
45751  TACAATCCCT TCTGTTGACA GGTCCCAAAC AAGGAGTTAC ATCACCTGGA
45801  TGCTTGGCTC TGTGATATGT CACATTGCCC AGTGCAGGCA GGGCACAGGC
45851  AGGAGAGTCA CATTACCTGG GTGCTTGGCC CAGTGCTATG TCACATTCCC
45901  TTACTTAGGG AGGGCACAGG TAGCTAAAGG GAGTCATATT ACCTAGGTAC
45951  TGGGCTAGAG CTATCAGTCT GTTGGTCATT CAGGGTGAGG GTTATGATCT
46001  CTTTAGACAT CCATGTAACA GAGCCCCAAC TCGTTCATCC TCAGAAGCAA
46051  AGGCCAGTGG AGACCAAATA AGGCTTGAAA GTTAAGTCCT CTAAAAAGGC
46101  AGAATGTCCT AGGCTTTGGC CTCTCCTCTT TAAGCACAAC CAGATCATAA
46151  CCTCTCTTTT GACTTCTCAC TATGCCGCTC ACTGGAACAA GGCTCCTGAC
46201  TCCACACCCT CGTCTTTTCA CCCCTGTACA TTTATTGGGC AGACCAAAGA
46251  CCCGGAGACC TGAACTCTGG GACAGTCTGG GGAGAAGCGA ATCCAAAAAG
46301  GGTTGGGGCA GCAACGGCTG AGCTCCCACA GATGCCCCGG CCCAAAACTG
46351  CAGGCCCAGG CTGTGGTCCA TGTGGCACTC AGTGCAGGTT CTGCAATGCC
46401  ACTGGAAGGC CTAGAAAATC CACTTCAGCT GCTGCTTCTT CCTTTTCTTA
46451  AGCGCTGAAA GGTATACACT TTCCCAAGCT TTTCAGATAC GGCCAGGCAC
46501  CCCCAACTGA AAAGGGGGAT AATCGCATTC TTCATCATGG TTTATCACCT
46551  CACCCAAGCA CCAGCCAGCC CCAGCCTCGC CAATCCAGCT GCCCCAGGCT
46601  GGAGGAGATC CTGTCAAGCC TCCTCTCTGA AACTCCACTC CTCCTCTTTT
46651  TTTTCTTCCT CAACCACTGG CATACTTTTG GTCATCCTAC TTTTCATCCC
46701  AATGGGGCAC CTGATACCAT AAAGACCTGG GAACCCCACT GATCACAGGT
46751  CCTACAGGGG TCAGGCAAAC ATACAGGAAG TCCAGAAGGA CAGTCATCTG
46801  CCACCTCCAG TATGAGAAGA ACAAAACCCC CGCACTCCAT GAAGAAGTCA
46851  GGAACCCAGT GAAGGCTGCA GGTCCAGCCC ACAACAACCT ATGTATGTGG
46901  GGTGCTCCAT GGCTCATACG TGTGTCCTGC AGGCTGGAGC CCAAGCGAGG
46951  GTTCGCAGTC TTTGTTCCAG ATGGGGAACT ACTACTTCAG CTTATTACCC
47001  AGGCTCCCTG ATGGGGAAAG GAAGGAAATG GCGAGAATCT TCAGCACAGC
47051  AAATCCCATC GTCCTGGCAG AGACAACTCC TGCCTGCATG CCACACCAGC
47101  AGCCGCAGGC TAGAGGCAGT CACATCAGGC CTCCCTCTTA CAACATTTTG
47151  TTTTGTTTTG TTTTTCGAAA GGCCCTAATC ACCTACTTTC GGTCATCCTC
47201  CTTGTCACAA TAGGGCACCT GACGCCATAG AGACTCGAGA AATGTACAGA
```

47251 TCTGGGGTCC TGAAGCGTCC AGGCGAAGGT GCTGGGAGTC AGGGAGGAAA
47301 GTCCTTTGAA GCCTCCAGAA GCAATGTAAC AAGACCCAGC ACTGCAGGAA
47351 GAAGTGGGGC ACCCAGTGGC AAATGCAGGC CTACGTGCAA GGTGCTCCCT
47401 GTCCCTCGCC TGCTCCCTGG AACCTGGAGC CGCACCTAGG ATGGGGGGTC
47451 TCCATTCCAG ACCCAGAATA AGCTCTTGGC TACCTCACCC AGCACCTGGT
47501 CATGCCAGAC AGAATCTTGG CTGGGGCTTT CTCATGGTCT GCAGTGCCTT
47551 CCACTTTGGG CCACCGTAAT AACACTTCTC CCACAGGGAC TCTGCAGTGT
47601 CCAGCTCACC AGCCTGGAGT TTCTCATCAC CCAGTGATTC CAAAAGAAAC
47651 AATCTACAAT GGCATGAGCC CAAGTGCCAG AAACAAACAA ACAAAATATC
47701 TGAGGTAAGT CCATTGGTTA CTTCGATTTT TCAAAGGTAA CATTTGTCCC
47751 TCCTTGAAAT CCTAAGAGTG CATGAACAGG CTATTCTAAT GGACGTGAAA
47801 TTCACATTAA AACTGATTGA CAGATGAATT CTGATTCCAA GGTACTGTTG
47851 TATTCTCAAA TTGCATCTGC TTACCCTGCC CCCCTCAAAA TGGAAGAGTG
47901 ATGACTATTT GTCATCTTAG CACTGTGGGG ATGCAGAGCC TTAGATGGAA
47951 GTGTGTTAAA AACAACATTC CATAAAGGGC TGTCACTTCC AATTTTCAAG
48001 CAAGGTGGGA AACCAACCGT GATGTATGAA ACACTGCTGG CCAATGTGCA
48051 CAATTAGTGA TACAAAATTG AATAATATAA ACTATTGTAA ATATGTGTGG
48101 ACATTGTGGC AATTTGGACA CCAAATAACA CTGCCAAATT CAGGAAGAGA
48151 AAGCTGTAAC CTCATGGTTG TTATGAAGCT TAACTTCTGG GTACTAGAAC
48201 CTATCAAAAT TTCTCCATGC CAATCTGTCC AAACAAACAT TGAGATGTTT
48251 ATTTCTATAA AATTCCTATA GAATCTTCCT GGGCATGACC CTTCTAGCCC
48301 CATCCTCACA GCACTGGTGC CAGAGGAACT GCACTTGCTC CTGCCTCCCC
48351 CTATATTTTT AGGAGTGGAA ATATTTAGTG ACAGGATGGA TTGGGAGGGT
48401 ACTGAGGCCT CCTGGGTGGG TGGGTCAGAT CTACTGCAGC CCAAGCCTTT
48451 TTATAATAAA TAACTTATAG GTAAATTAGA AAAGAATAAA AAATGCAACT
48501 CATTCTCTTT CTTTGCTCAC CACAAAGCAG TAAACACAGC ATCTAGAGAT
48551 GCTGTTGGAG CCAGGCTTGT CCTGGGAAAG TAAGAAGTGC TTATCAGGAG
48601 CCCTGGGATT GAGGGCAGGT GAGATGGGTT CCCAGGAAGA TACAGCCAGG
48651 GTCAGGTCTG GCCCACTCAC ACTGGAAGGG GCCTTCTGAA GGCCAGGAAA
48701 AATGGTCCAG GTCACTGAAA CTCAAGCGGC CCATCTGAGC CAAAGATCTG
48751 TCCAGGTCGC AGTGAATCTC ATCAGCACTG CCAATAGGGG GTCTGATCAG
48801 CCCTGAAGAG CTCCATCAAT GGAGGCAGAT GGCTGGTGGA TGGGTCAGGA
48851 GAGCTGCCTA CTGCCAATGT GGGAGTCCAC TCGGTGTGGG TTTACTGCGC
48901 CTCTCAGTGG CCTGGATAAC CTGCTGAGGC TGCAGCTTCT TCCCGTTTTT
48951 GAGCAAACGG GGACATGTAT CATTCCCGAA GGTTTTCAGA TAATCCCTGG
49001 TGACCCCTGG CAGGGGATGG TTATCTTGGC GATCCCCAGC CAGGCTTATT
49051 GACTTGCCCC CAGGCAACAC CCAGCAATCC CGCGCCCACC CAGGCACTTA
49101 AGCATTTGTG GTAAGCCTCT TTCTGGAAGC TCGCCTTCTG CTTGCCCTTT
49151 GTCTTCCCCA CCCCAGTACT TCTGGCCATT CTCATTGTCA TCACAATGAG
49201 ACAACTGGCT CCTGGAGACT CAGAAACTGC CATGCAGACC TTGAGTTTTT
49251 CCTAGGCCCG GCCAACAGGG TGGAGAATCT TCCATCTCCT AAAGGAGCAG
49301 AACAGACCAG GCATGTAGGA GTTGTGCCCC TGTGCAGGCT GCAGGAGCAG
49351 CTCACAACAA GGCCTGAACA TGGGGTGCTC CCTTGTTTGT CCTTTTTTCC
49401 TGGAGGCTTG GGCTTGCCTG GCCAAGGTTG CCCTTCAGGA CAGGGAATCA
49451 GAGCTTTTTG TGGCTGCCTT ACCCAGCCAG GGCAGCACCA GAGAGAATCG
49501 TGGCTGAGGC TTTCCTGTGG GCTGCAGTGC CTGCCCCTTT AGGGTTTCCA
49551 CAAAAACACC TTCCCTGCAA GTAATCCACT GTGTCCTGAT TACCAGACGA
49601 GGCTTCCTCA TCATTTGGTG ATTCCAAAGA AAATAATCTA CAGCGGAATG
49651 ACCAAGTTCT AAAACAGGAG CACACAAATA ATCTGTGGAA AGTCTGTTTG
49701 TCATCTAGAT GTTTCGAATA TACACATTTT CCCTTCTTGT TATATTCAGT
49751 CGTGCATGAA ACTGATATTA TAATAGATGT GCAATACACT AAAGCTGATT
49801 AAAGGGTGAA TTCTTATTCT AAGGTACTTT GTGGTCTCAA ATTTGTCTGT
49851 TCCCACCCCC AGGACTCCCT TAAAACAGAA TAGTTATCAC TGAGAGCAGA
49901 GGTAGAAAGA AACTAGCTAG GCAGATAGAG CAAAGAGTAC TCAGCGTAAC
49951 ATCCCTTCTA ATGAAAAGCA GCCCAAAAAA TCACATCTCT TTAACAAAGA
50001 GCAACCTGTA AGTTCGGGCT GCAATCATAG ATAAGTAAGA TGGAAGCTTG
50051 TATGGGCAGG GATGGCTGCA GCTTCATGGA TAGAAATGTC CAGCTTGGGC
50101 TAGATACATC CAACATGGGG GCTCCACTCC TCTTTGTAGC ACACGCACCA
50151 TAGGAAAGAG ATAAGCAACT TGGAGTAGCT CAAAAGTCAC GGAGCCTCAG
50201 TGTCCCTTCT GTGGAGCCCA GAACCTGATG CAGGTCTAAG TCCTGTTGTA
50251 TGAACATGTC CTGACCCTGG CGGCCCTGGT GGTGGTGCAG CATAGGAAGT
50301 ATAAGGGATG AGGTCTAGTC ATGGGCCATG GAGCCTTTCT CATTAATCTT
50351 GGCTGTCTGC CTTCTAGGGA ATATAATCAA CACTAATAAA GGAGGAAGGT

```
50401  GAGCAGCTGG CGCTGTCGCT TTGAGGGAGG ATGGCGATGT GAAAGTCAGT
50451  GACCACCGTG GGGAGGACAC TCCCTGGCTC CATCCTCTGC ATCTTAGATT
50501  TATTGGGACA GTTTGATACA CAGAGAAGGA GGAGACCCAT CCCAATGGAG
50551  GGTTTGATTA GATGAATATA ATCAATGATA AATTCCTAGA GGAGGGACTT
50601  TTTATAATCA ACTCTGAGAA CAGGTTGGAG CTACATGGGA TTGGAGGGGA
50651  GGGTGGAGCC CCTTAAAAGA AAAGCCCCAG AGACTGCCCC TGCCCTCTCT
50701  CTCCCCCACA AGTTCCATTT ATTATCTTCC ACCCAGGAGC TGTCAGAATC
50751  CTGCCCTTCC GTCTCCAGAT CAAAGTCCTT CAGGAAATGC AACTACTTCA
50801  GTGACAAGAG ATAATTATCA TCTTCTGACA GAGGAGGAAT TTGGGGTTTG
50851  GTCCCAGTCC ATGAAGTGGC ACAGTCAGAA TAAAAGGTGA GAGCTTAGGA
50901  GATTAGCGGA GGGTAGAAGA ACACTCTGTC TTGTGACCAG CTTCAGAGAG
50951  CCTGGGGCCA TGGCTTCCTG GTCAACATTA GGCCCTGCTG CATGGTGACC
51001  CCTGGGCAGG CAGTGGGAAG CCTGAGGTGT GGCTCCTGGT GGCCTCACAA
51051  CTGCCACTCT TTCCTGAAGC TCCTATTTGT TCTGTCAGCT AAGCCCCCAT
51101  CCCAGTAGGC CAGCAACACA CTCAAGACCA AGAACAGGCC ATGGTGAATC
51151  TCAGGGCCAC TGAGTGCCTG GGCTGGCAGG GGCAGAGTTC CTCAGGGCTC
51201  AGTGACATTT GGACTGAGCA TGGGCTTTGG AGTCATACAG CTACACTGAG
51251  CTCCCAGCTT CACCGTGATC AGCCCTGTGT CTGGGACAGG GGCCTCACTG
51301  TTCTGGAACT TGAGACGCCA TAGTCATAAA TTTAACACAC ACTTCTAACT
51351  GCTTTTTCTT TTTTATCTGT CTTTCTCTAT AATCACCATG TACTACTGGT
51401  CTCTTGTGTT TATTTAAATT AATAAACATG TTACACAGTG TGTATTATTC
51451  TTCCTCATGA GTTCTTTACT ATATTCTATG ATTCCACCCA TATGAGGTAC
51501  TTATGTAATT TCATTCATAG AAACTCAAAG TAGAAAAGCA GTTAGTTCTT
51551  AGAGGACAAA AGGAAGGTAA AGGGGGATTG TTGTTTAACA GGAACAGAGT
51601  TTGAGTTTTG CAAAATGAAT AAAATTCCCT GTGAATGTGG ATGATGGTTG
51651  CAGAACAATG TGTGATTAAT TCCTCTGACC TGCACTTTTA AAAATTGTTA
51701  AAATGGTTAA TTTTATGTAT ATTTTACCAC AATGTTAAAA AGGACTTTTT
51751  AAAATGAACG GACTATAGAT ATCTGCAACA GCATAAATAA ATATCACAAA
51801  TATAATCTTA CATTTAAAAA TTGATGTAAA AGTATCCATA CTCTGTAATT
51851  TCTTGTATTT AAATCCAAAA ATCAAAACTG AGGTTCTGGC TTCCACTAAT
51901  GATGAAGTAG CTAGTTTAAC TAACAATCTC ACAGAGAAAA ATGATGAATC
51951  CCAGGTAAAA CATTATATGT TATTATAGAA ACACTTCTAT ATATAATAGA
52001  TATATGAAAT ATGTGTGTAT AAAAACTGAA TGCATATTTC AGCTGTGCCC
52051  TCCATAGAAG AGAGAAGTAT TGAAGTTAGA AGCCAGCCCA ATTAACACCC
52101  TCTTTAAAGA CAACACTCTT CAAACGGACA AAACAGAATC CAGAGTCTCT
52151  TTAACTCTTC TATACAGTCT CTAGTGCACA ATTTTCCAAT TCAGGAGATG
52201  CGTGAAAACA CATGAAAATA TAATACATAC ACAAGATAAA AAGCAGGCAG
52251  TAGACATCTC CAAGATGTCC AAGACATAAT CAGCAGACAA GAATTTGAAG
52301  GCAGCTATTA TAAGCATGCT CATGGGGGCA AAGGAAAATA TTCTCATAAA
52351  TGAACAGATG TGGAACATCA GCAGAGAAAT GAAAAATGAC CACATAGAAA
52401  AATAATAAAA ATAATTGTGA GCTTTTCTAT ATATCAGAAA CAGAAGACAT
52451  AGCAATATAA TTTAACCAAT CTGAGGATAG AAGTAAAAAT AGTTTAAAAG
52501  AAAATGAACA GAGCCTTAGA CCTATCTGTG GGATGATTGA TTCTGAGAAG
52551  AAGAGAGAGA CAGAAAAAAT TAAATGGGGT AGAAAAGCAA ATCAACAAAA
52601  TTTAAAGAAA ATAAACCGAG GCTTAGAGAC CCATGGAATC ATTCAGTCTG
52651  AGAAGGAGAG GAGAGAGTCA GAAGAATTAA AAGGGTTACA AAAATAAATC
52701  AACAACTAAT ATCTGAAAAC TTTCAAAAAT TGTTCAAAAA CCTAATTCTT
52751  TTTTAATCTA AAGATCCACA AACCCCCCAC AAAAATACAA ATAAAACCAT
52801  ACCAAGGCCA TATTGTGATT TAAGAAACTA GCAGACAGGA CTTTCAATTG
52851  ACTTGATATG ATTTATTATT TTTACTACTT ATAAGAATGG AAATAAGTTC
52901  TCCTTAGTTT TTTTCTTGGA GAAAGTCTGA CATGTGAGGC ACAGATGAGT
52951  TATTAAAGGC AGATGACTTT CCAGCCTTGT CTTAAATGTT CCATTCTTTA
53001  CCTTAGAAAT TATTTAAATT TGTGTCTTCC AAATACTGTA GTAATATTGA
53051  TGCTCCAAAG AGATGTCCCA CGGAGATTCT GCTCTTGTGT GTCCACCCTG
53101  CAGGGAGCTG AGGCAGTTTC TTATGACAGT TTCAGAAGCG AGTAGTCGTG
53151  CAGTACTTAA TCTAAAAAAC TTAATGGAAA CATGAATTAA GAGAATGATC
53201  ACTGTTTAGT TCTATCAGCA AACTATTAAA AGTGATCCAA AGGAGGTATT
53251  TATAAAGAGA TATTAAAAGA TTTTTCAAGG GAGCCTTATT CAGGGCAGAA
53301  ACGCAGACAC TATCGCTGAC CTCACCACAG AAAATACCCT CATGGGTTGG
53351  GAGGGACCAA GGGACGCTCT GGTCCTGCTG ACCTGCATTA ATCACAGCCA
53401  GGAGGTCCAC ACTAGTACCA TGAGGCCTGG GAAGCAGCCT GCGTGGGGTC
53451  AGAGAAGTGG TGGATGTGGC TCCCAAAGTG GCTTACGGGG TCCCTTCCCT
53501  GTGGCTGTTT CCTTACTGGA TGCAGCAGGG TCAGGCCCTT CCCCTGTGAC
```

```
53551  GTTTTCTCCT CTTTATCACA GTGGCGGGAG CGTCCCCGTG AGAGGCCCGA
53601  CCCAGGTGTG GGCCACGCTG CGAGCCCGAG GACCAAGCGG CACTCCTGGG
53651  ATGCAGAGGA GGATTTGTGA CAGCTTAGGG AACAGAAAAA ATGGTTTCGA
53701  AAAGGCTAAT GGCAGGTGAC TAAGGACACG ATGTTTTCAT TACTGGCAGT
53751  GAACTGACGG TTTCATACAC TAACAAGGGG GCTTCTCGAG GGGATCCCAA
53801  GGAGCCCAAG AACTGCCAGG TCGCCCACCA TTACCCTACG CCTAAGGGAC
53851  AGGCTGCACT GAGCATGTCT GAAACGGTAG GCCCGTTAGC CCCACCCCTA
53901  GGAACGGGTG CACTCCGCAT GTGTAAAAGG GCAGGACCTT TACCCCACCG
53951  CTAGGGACGG GCTGCACTAC GCATGTCTGA AAGGGCGTGA CAAGAGGGAG
54001  GAGCAAGAAG GGGCGGGGTG GAGGGGGAGG GGGGCAAGAA AAGGGGCGGG
54051  GTGCGCCCAA CATCCGGCGG AGAAGTATTA CCATGGCAAC CCTCCCGCGC
54101  AGGCCATAAG AGGCAAATGA ACCTTTGTTG GTTGGCGGGA AATCGAGACC
54151  CTGGCAAGGG GGCTTCTCCC TTGGAGAAGC CTGAGAACTG CCAGGTCCGC
54201  GGCCGTTAAC CCGCCTCTAG GGACGGCCGC ACTGCGCATG TCTGAAAGGG
54251  AACTAGAAAG GGAGGAGCGA GAGAGGGTGG GGCTGAGGAG GAGGCTGTGT
54301  GAGAAAAGGG GCGGGGCGCG CCCAAAGTCT GGCGGAAGAG CGTTACCCTG
54351  GCAACCCTCC CGCGGAGGCC GAGAGAGGCC ACCGGCCCTT TGTTGATCTG
54401  CGAGAAATCA AACTACGAAC ACGACAAGCA TTAGCCTGCA GCTCGAGGAG
54451  ACAAGGTGTC ACAATTACAA GGGGAAACTA GCCGCCCTAG CTCCACTGTC
54501  TCCCCAGCAC GAGAGATTTG AGAACAGAAA GGCTTCCCTC CGCAGGGCGA
54551  AACTGCTGGG CTGCCTGGAA GGGCGAGGCA GGGAGCGGAA CCGTCTTCAG
54601  GAAATTTCGG GAGTTCCGGG GTCAGGTCCA CTCCCCGGCT GTTGTTGTGT
54651  TGTTGGCAGG GCAGAGGGTC TAGGATGCCA GCCTGCTCCG GGCTGCGCTG
54701  TGCGCCTATC CCAGGGCGGG GGGATGCGGG GCGACACCCG CCTCCCGGTG
54751  CATCCAGGAG TTGTAGTCTC TTCACCGGTT CCCCACTGTG GGTGGTGGGG
54801  CTGCAGGAGG ACAATTCAAA TTGAGATAGG AGCGGAGGCG GAGCGCGGCC
54851  GTGCAGGGAG GGGCAGGGCG GTGTAGGCGG CTTCATTTAC CAAGCTTTGC
54901  TGGCCATTGT TTCCATGCCA AACCCTTGCC AAGGGGATTG TCAGGAGAGG
54951  AACTTGAAGG GGAGGCGTGG GCTGGCCAGT GAGGAGGGTG TGTTTTTGCG
55001  AAGTGCGCCC CGTCTTTGCC GAAATTAGGA GTGTCTGGTC CTCACTCACG
55051  CGGCTCTCTG CTGCTCAGGT CGATTTTCTC TCCCACCCTC ACCCAGGCTC
55101  TTTCCACAGC ATCACCCCTG CCCCAGCCCC TGGAGCCACC TGCTTTCCTA
55151  AGTGGTTTTG GAAACTGGGC TGAGGTCCCA AAGGGTGTCC ACTGTGCTGT
55201  TGCTCTCCGC TCTGTCCAAG CAAAGCACAA GCTCAGCCGA CTTTGAAAGA
55251  CACCCACCGC CTGGCCTGGG AATGCACAAG TTCAGAGCTT TGCAAGAAGT
55301  GATCATGGGC TATGGCTTTG TGAAAATGTC ACCCTCACCA GTGCCTTTTT
55351  CGCGGACGTG GACGTGGAGG AATGAGGGAG GGTAATCACT GGGCTACCAA
55401  GGTACTGCTA AGAGCAGAAG AGAAAATCCC AGTTTTCAGC CATGTGTCTG
55451  GTTTGACATT TCACCAACCC ATTTAAGTGT GCAGGCCCCC AAATATCTAC
55501  CTAAAGATTA TGATAGTTTA GGCATTTTAC ACTTGAAATT ATTGACCTCA
55551  TATCCACTGA AGCCTGACTG GCCAGTGTCT CAAAGACACA GATGATGACC
55601  TGATCCCTCA GGAACAGCTG GTGCTCCAGC TTTGTGGAGG TGAATTTCAA
55651  GGTATGGATT ACTTGGGGGG TCGTTGAAAC CTGTCAGGCT CAGGAACAGA
55701  TGGTGCTCCA GCTTTGTGGA GATGAATTTC AAGGTATGGA GCACTTGGGG
55751  GGTCTTTGAA ACCTGTCAGG TTTCATATCT CTGCTTTGTG TGAAAAGATC
55801  ATCACCTACA GTAGTCAGGG ATGTGCCTTT ACTTACTGGT GCTCATCTGT
55851  TGAAACTTTA TATCTAGACA ATGGAAACAT TGAGGAGCAT TTCTGCTTTC
55901  ATGTAGCCTC TTAATAATTG ACGCCCTAAA GCCCTGTGTC CTCAGGGAGA
55951  GTTCCTTTGA TTCCTGGGTG GTACCAGGTT TCATGCTGTT AAATCTGTAA
56001  AAACCTGCGC GTTAATCTCC ATGAATATAA GACCTTGTTC TTTCTTAAAT
56051  GCCCCAATTT TTTCTTCTCT TGTCTTATAT TCTGAGCAGG ATTTCAAATA
56101  CTGTATGTAA GGAAGTAGTG AGAGTGGGCA TCTTTATCTT AAAATAAATC
56151  TTAGAAAAGA TTTCAACATT TCACCACTGA CAATGTTAGC TATGGGCTTG
56201  TCCTATAGCT AATAAAGAAC GCATCTCTTT ATTTTGAGGT ATATTCTTTC
56251  TATACCTAAT TTGCTATAAA TTTTGTTAGG AATGGATTTT AAATTTTGTC
56301  AAAATAATTT TAGGCATGCA TAAAAAGTCA TGATTTTTAA TCTTTTTGTT
56351  GTGTAAATAA GGTGTATAGC ATTTATTGAT TTCCACATAT TAAAATATTA
56401  TTGCATCCCA GGAATAAATC CAACTTGATC ATAACATCCA ACTGTGTTGC
56451  CCAGGCTAGT CTCAAACTCC TGGATTCAAG AGCCCCTCTC CTCTCAGCCT
56501  ACGAAAGTGC TGGGATTAAA GGTGAGATCC ACTATGCCTG GCCATATATA
56551  TATATAATTT GTAAATAAAA ATAACCTTAT ATACAAGGAT AAATTCAAAT
56601  GTCCACGGTG AGGTCTGGGC TTCAGCATAA GGAGGAAGTC TTGCCTGAAA
56651  AGGGCTGCAG CTTGGAACTT TTTACCCTGT CGTCATGTGG CTATGAGTTG
```

```
56701  GTTCACATCT TCTGTCATTC AGGACCCGAA GGGGTGGGAC CTGGGGCCCT
56751  ATTATCACTG GTGCTGGGGT AAAAACTGTC TTAAAACTAT CATTTTAATG
56801  CTTAGCAATG TTAATTTTTA GTGAGAAACA AGATTACTTA ATTTAATATA
56851  ACCAGATTTT AAGTTACTAA AAAAAAACCC TAAAATTATG ACACAGGTAT
56901  TTCCTCTAAT GTTTTTTTGG GGGGTTTCAA GTACCTATGT CATATACTGA
56951  AACCTACTGT TCTGTAAGCC CTACCCTTAA AACAATCTTG TTGTTATTGT
57001  GAACAGTTAC TTAGTGTAAA TCCTACCCTT AGGCAAATTT ATATGGTGAT
57051  TTCAATTGTG CTTCACATTC CTTTCCTGTG TTAAGTGTCT GGGTTTAGGG
57101  GTTAACAGTG GGAGGATCCA CTCATCTTCA GCCATCTGAG ACATAGCTTC
57151  TATTCATAAG TCCATCTTAA ATGTTCCTTT CTGAGAAACT TGATTTGTCA
57201  GCCTCATTCT TCAACCTTTC AACTCCCTTG GCTTTTAAAG GCAGGTTTAC
57251  ATATACCTAC TCACAACAAA ACACCCTCAT ATATATGGGC TGTCAATTTC
57301  TATAACATTT TTATGTGGTT CAAGACTGTA ATGTGTAGCA CATGTAGTTT
57351  TGTATATGGA TAGTATATTT TATATAGTAT ATTTTATATA GCTACTTTAT
57401  ATTACACATC ACTAAAATAC ATGTTCAGTA AGTGCTCACT TAACGTCATT
57451  GATAGGTCCT TATAAACTGA CTTTAAGTAA AACAAAATAC TGTGTGCCAT
57501  GGAAAATTAA CTTGTGTATA TCAATTAGCC AATGGGAAAA TTGGGTTTAT
57551  TATATAGTAT ATTGTGTTAC TTAAAGTCAC AGTTTCCCAG AATCTATCAA
57601  AAAAGTGAGA ACATACTGTC ATTAGTATTA TGTAGTACAT TACAGAATTA
57651  CACTGTTATG GTATGTTATT GTAGTCTTAG CAGTTGGTAG TATAATGTGT
57701  TTCAGTTTCC CCCAAGGTCA CAGAATTATC CAGGCCAACC AATAACACCT
57751  CCTGTGGGAA CCAGGAGCAT CTCACCCTCT TGATACTACA AAGCCTTCCC
57801  CGACACCTCC TGTTTGTTCT CTCTGCTCCC AGGTGCAATG GCTGTGTGGG
57851  TCTGTATACC TTACATAGCT TTCTCCTTCC ATGATTATAT GTAATGAATA
57901  ACTGCTGTCA ATCTCATCTG TCCAGTGATT GGTGCCATGG TTTTAACTAT
57951  TCCAGTAGCA TTAGGGTGGT AATTTCTCCC TCACCAATGG GGTAAAGGGG
58001  AGGCTAATCA AACAATTCAC AACACTAACT GGATTAATCA CCCATGACGG
58051  AGGACATCTG CTCAACTTTA ACTGCTTTTG GCCTACTGGT TTCATGATAC
58101  ATTAAAAGTC ATCTCTGTCA GAGCCATCAG TTTGTGGTGG GCTTTTGCTT
58151  TGGTCTAAAT AACAATTTGT GGCCTTTATC ATGATTTGCC TTCCCTGCCC
58201  ACACTAAAGC ACACATTACC TAGACATAAA TATTCAATGG ACTCTTCTCC
58251  CGCGTGACGT GTCATCATGT TTAGCCTGTG TTTGGCAGGC AGTTTGCAAG
58301  ACACTTGCGT GTCAAGGCAG TGAAAACACA GCACNNNNNN NNNNNNNNNN
58351  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
58401  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNGGCAGG GAAAACACAG
58451  CACCTTAATC AGTCAGCACT TCAGGCGGGA TTACCACAAG AACGGCCATA
58501  TACCTGTGGT CGCTTCATCT GGTCTGCTTA CTATTACTAA ATGCCTGGGT
58551  AGTCATCTGA ACGTTGTTTA TTATTGCACA CTCTCAGGAC AGACCCAGGG
58601  ATGGTCTTCT GTAAAATTGC AAAAACAAAA GGGACTTTTA CTTTGGAGAG
58651  AATTATGCTT GTCATTCAGT TGCTCTAAGA AGTGCTACTG TCATGCAAGC
58701  AAGAACAGTG GCCTTGTCTT TTCTGAGCTG CTGCTTCCTT TGCTGCTGTC
58751  ACGTGCAATC TCATGATGAG GTGTGCATTC CTTGTGCCTT ATGGCTTAGG
58801  CACTGATGAG TCATAAAAAG GAGAAAGAGA ATTTAATATT AGTCCCTCCC
58851  AAAACATTTG GGGTAATTCT CTCACTGTAA AACCCCTACT CATCATGTGA
58901  ACTTTTAGCA CTGCTGCTTA CCAGTATGCC AATGGTGCAA TCTTAGGATC
58951  AGAAGTTTGA TGGACTCAAA AAAAAAAAAA AGTACCATAA CGAATATGGG
59001  TGCTAATCCC AGTGAGATGG AGAGGCAGTA AACACCTTCT TCCAAAGAGA
59051  AGACAGATTG CAAATAAACA AATAATGATA TAACTCAAGA AACTAGAAGA
59101  GGGAGAACTT ATCCCAAAGT TAGCAGATGA TGAAATTAAT TAAAAAAACA
59151  GAGCAGAAAT AAATACAATG AACACTAGAA AAAAAAAACA GAAAATGTAG
59201  AAAAAATGTA AGAACTTGTT TTTTTTTTGA AAACATAAAA TCTTCCTATC
59251  TTTAGCTAGG CTAAGAAATA AATGCTCAAA TAAATAAAAT TAGAATTAAA
59301  TGGAAGATAT TACAACTGGC AAAACAGACA TACAAAAGGT CATAAGATTG
59351  CTGTGTACAT AGACTCCTAT TTACAATGAT ATCCAAACAA ATTGTGTAAG
59401  GGAGAAGACA TGAATAAATT TCTAGACATG TACCACCTAC CAAGACTGAG
59451  CCATAAAGAA ATAGAAAACA TGAATAGATC AGTAATGAGT AAAGAGTTTA
59501  GATCTGATGA CTTTACTCCT GAATTTTACC AAACATTTAA AAAGAACCAA
59551  ATCTTTGAAA AAAAATTGAA GAAACAGGAA TACTTCCAAA CTCATTTTAC
59601  GAAGCCAGCA TTACCCTGAT ACAAAAACCA GAGATAGACA TTACAAAAAA
59651  AAAAATTACA CGCCAATATC CCTGATGAAC ATAGATGCAA AATCATCCAC
59701  AACATAGTTG CAAATGAAAT TCAAAAGCAC TTTAAAAGGA TAATTTATTT
59751  GATGTGGGTG CCCCCCATCC CACCTTTTCT TAAAAAATGA AAAAAAAATA
59801  GCATGGCATG GTCATACATG TCTGTGGTCC TACCTACTTG GGAGGCAGAG
```

```
59851  GTGGGAGGGA CACTTGATCC TGGGAGGTAG AGGTTGCTGT GAGCTGAGAT
59901  CAGGCCACTG CACTGCAGCC TGGGTGACAG AACAAGATCC TCTCTCTCTT
59951  TCCCCCACTT TGTGTGTGTG TGTTGCACAC ACAAAGGGGA GGGGTGAGGG
60001  TGTGGATGTG TGTGTATTAT TCAAAATGAA AACAATGACA ATTATTTTTG
60051  TGTACATTTA TATTTTTTGG AGAAAGAGTA TCACTCTGTC ACCCAGGCTC
60101  AAGTACAGTG GTGTGATCTT GGCTCACGGC AACCTGTGCC CCCTAGGTTT
60151  AAGCGATTCT TCTGCCTCAG CCTCCCAAGT AGCTGGGATA CAGGTGCCCA
60201  CCATCTTGCC AAGCTAATTT TTGTATTTTT AGTACAGACA TTGTTTCGCC
60251  ATGTTGGCCA GGCTGGTTTT AAACTCCTGA GCTCAAGTGA TCACCCACTT
60301  CAGCCTCCCA AAGTGCTGGG ATAACAGGCG TGAGCCACCG TGCCTGACCA
60351  ATGACAATTA TTTTTAAATT TTAGATTTTA CAATCTTTCT CTCCTCTTGA
60401  ATTTGAGGCA GCCTGAGCTT TGAAAACTGG CAATCTTCTA TTGCAGCAAT
60451  GTGCAGTAGA AATAAAATGT GAGCCACGTG TGTCAAAGTT TTCTAGTAGC
60501  AGCATAAAGA AAAAGAAATA TGAGTGAAAT TGATTTTAAT AACTTAGCCT
60551  AATATATCCA AAATATTTTA ACATATAATT AGAATGAAAT CATTACATAT
60601  ATATAGATAG ATATAGATAT AGATATATAT GTAACTAAAT CTCTAAAATA
60651  CTTTCTGATA TTTTTCCCTA GCATATCTCA GTACATACTG TGTACAATTT
60701  GGGTGCTCAG TAGCCTGTTG TGGCCAGTGG CTGCCACATT GCATGTGCAT
60751  CTCTGAGGGC TCTTGACTTT TCTGACCTTC GGGAGGTAAA GGGCCTGAAT
60801  TTTCCTTTTC TGCCAGATAG GAGTGAATGC CTCTTCTCTG CCAATATCAC
60851  TCCTGTTTCA AGGGTAAGAG AGGTGGTGCA CTGAGAGACA GGAGGGACCT
60901  ACAGGAAACA AGTGTCCACA GATAGACCAC TGCTGTCCGC TGCTGCTTGG
60951  TGTGGACTCA TCACTCTTCC AGAAATCAGA CAGAAGTCCA AAAAATGGGG
61001  ACCCAGAAGG GGGAAACCTC ATGTTTTTAG ATCTGTCCAT AGCCTTAATC
61051  TCCAGAATTT AGATGTCAAG AGACTAGATT AAAGGCAAAC CTTTTATCTT
61101  GCAATTTGGC TTTGGCAAAT TAAAATAGAA ATAGAAATGT GTCACTATAA
61151  AAATCAATTA TATACAAAGG AAGGCAATAA GAGCAGAAAA GAGGAGAAAA
61201  TACCTACGAG AAACAAATAG AACTATTAAC AAAAAGGAAA TATTCCATTC
61251  TTTTCAGAAT TAAAATGAAT ATATACATGG AGTGAAATAT ACACTGGAAA
61301  TATGTAAATT GGCTAAAGAG ATTTTAAAAA AACAAGATTT ATTTTCTGTT
61351  GTCTATAAGA AACTCAATTT ACATCTAAGC ACACAGATAG GCTAAAAGTG
61401  CCAGTATGAA AAATACCTTC TAGGAAAACT GCAATCAAAT GACAGCAACG
61451  TGGATCATAA TTATGCAAAA TACACATTAA GTCAGAACTG AAACAACAGA
61501  CAAAGAAAGA TGTTGTATAA TGATAAAACT GTCAATTCAC TGGGGAAGCT
61551  GTGTTAATAA TAAATATGTG CACACTTCAC ATCAGGGTTC CCAAATGTAT
61601  AAAGTTAACA TTGACACAAA TGAAGAAGGA AATGGCTATG CAAAAATAGT
61651  AAGAGACATA ATTACCCCAC TACCCACTAT CAGTAATGAA TAATAAAGCC
61701  AGACAGAAAG TTAACTTGAG AACAGAGAAT TTGAATAACA CTGCAAACTC
61751  TAAACCTAAC AGACATATAG AAAACACTAG TCACAGTGAA TAGAAGTGAA
61801  AAAACAAAAG AAACAAACAC TCCACACAGC AATATCAGAA TATACAATTT
61851  TTTCAATAGG TCATGAAACA TTCTCCTGGG TTGATGACCT ACTAGGACAC
61901  AAAACAAGTT TTGCTAAATT TTAAAATGGT AAAATATGGG CCAGGGATGG
61951  TGGCTCATGT CTATCATTCC AGCATGTTGG GAGGCTGAAT TGGGAGGATT
62001  GAGTGAGTTT AGGAGTTCAT GGCCAGCCTG GGCAACATAA GGAGACCTTG
62051  TCTTTACAAA ATATAAAATT AAAAAATTAA CTGGGCATGA TTACACGTGC
62101  CTGTGTGTCC AGCCACTCAG CAGGCTGAGG TGGGAGGATT GCTTGAGCCT
62151  GGGAGATCAA GGCTGTGGTT AGCCATAATT GAGTCACTGT GCTTCAGCCT
62201  GAGTAACATA GCAAATCTCT GTCCCAAAAG AGATTAAAAT ATTACAAACT
62251  ATCATTTTTG ATTAAAAGGG AATACAATGA GAAATCAATA GCAGAAAAAA
62301  TACTGGAAAA TCTACAAATA TGTGGAAATT AAACAACCCA CTCTTCAGCA
62351  TGCTCTCGTT AAGGGTCGGA AGACAATATT GTGAAGATGT TCACACTACC
62401  CAAAATTATC TACAGATTCA ATGTGATCGC TGTCAAATTT TAACTGTCAT
62451  TTCATTTGCA GAAATACAAA AAAAATTCTA AAGCTTATAT GGAATCTAAA
62501  GTGAGTATCA AGAGCCAAAC AACTTTCTAA AAGCATAATA TTGGAGCTAT
62551  GACACTCCTT GACTTCCTAA TGTATTACAA AACTACAGTA ACCAAACTAT
62601  TTGGTACTGA CATAAAGGCA GACAGACAGA CCAATGGAAC AGAATAGATC
62651  ACAGGAATAA ACTGTCATAT ATATGGCCAA ATGAGGAGTT ATTTTTATAT
62701  CCATATTCAT TGCAGCATTA TTCACAACAG CTGATAGGTG GAAGGAACCC
62751  AAATGTCCCT CAGTGAATGA GTGGATAAAG GCAATTTGGA ATATACAAAT
62801  AATGGAATAT TATTCAGTTT TTTAAAAGCA GGAGATCTGA TTATTTTTAC
62851  ACTAAGAATA AATCTTGAGG ACATTATGTA AATGAAATAA ACCAGTCACA
62901  AAAGGACAGA CACTGTTTGA CTCCAGTTAA ATAAAATATC TAATGTAGTT
62951  AAACTCTTAG AAACAGAAAG TAGAATAGTA TCAGTCAGAG CCTTAGGGGT
```

63001 GGAAATAAAA GGGTAGTTGT TGTTTCATAG GTATTGAATT TTAGTTTTAC
63051 AACATAAAAT CATTTTAGCG ATATGTTGCA TAGCAATGTG AATATATTTA
63101 ATATTATTTA ACTATGTACT TAATATATTT AAGATGGTAC ATTTTATGTG
63151 TTTTGAGTAC ATTAAAAATG AAAAACTTTC TAAAGAGATA CATATTTATA
63201 ACCTTTTTCA AAAATTACCT CCAAATCATA AAAATGTCAG AAAAACAATA
63251 AAGAGGCCAG GTGCAGTGGC TCATCCATGT AATTGAAATA CAACAGGAGG
63301 CTGAGGATGG AGAATAGCTT GAGGCCAAAA GTTGGAGACC AGCCTGGGCA
63351 ACATAATAAG ACCTCATCTA CAAATCACAA GCAAAAGGAA ACTGGCAAAT
63401 TAAAAAATAT GTGAGACTTA AAACAGCAGA CTCTTGACGG CTCAAAAAAT
63451 TTAATATTAT TAAGATGTCA ATACTACTCA CAGTGAAACA CAAATTCAAA
63501 GTATTTTCTA TCAAAATCCC AATGTTACGA TTTTTTTAGA AATATTATTT
63551 TAAGTCCTAA AATTCTTACG GAATATCAAG GGACAATGAG TAGCCAAAGA
63601 AGCTTTAGAG AACGAAGTTA GAGGTGTCAC ACTTCCTGAT TTCCAAACAG
63651 ATTACAAAGC TATAGAAATA CAACCAGAAA GACAAATAGA TGATGGAACA
63701 GAATAGAGAA CCCAGATATA GACCATCATG AATATCATCA GATAATCTTC
63751 AAACAAGTTG CCATTACCAA ACAACAGGGA AATAACAGAC CCTTTAACAA
63801 ATAGTGTTCT AAAAGTGAAC ATCAAAATGG AAGAAAATGA AATTGGACTT
63851 CTGACTTGAA CCATATACAA AAATATCTTA AATAAATTAA ACAAATGTAA
63901 GTAAGATAGC TATAAAACTC TTAAAATATG AGGTAAAAAT CATGACATTT
63951 GTCTTGGTAA TTTTTTAAAA TATGACATTA AAAGCAGAAG TAACAAGAAA
64001 AAAAGCAGAA AAATGGGACT ACCTCAAATG TAGTAAGCTT TCTGGACATA
64051 AAGGAAACAT TTAATGTCAC GTAAGAAATG AGAAAAAAAT TACAAATGAT
64101 ATATTTGATA GAAGTTAATA ACCAGAATGT ATAAACAACT TTAAAACTCA
64151 ACAAAAAAAA CTGAACAACC CTATTTAAAA ATGGGCAAAA CTCTCAACAG
64201 ATGTTTCTAC AAAAGAGATA TACAAATGGC CAAGAGGAAT TTGAAAGGAT
64251 GGTCAAATTC ACGAATCTTT AGAGAAATGA AAAGCAAAAT CCCAATGAGA
64301 TATTACTTCA CACTCATTAG GATGGCCACT ATCAAACGAG AGAAAATAAT
64351 AAATATTTTC AAGGATGTAG ATAAATTGAT ATACTTGTGC ACTGAGTGGT
64401 GGAAAAATAA TAATGCAGCC ATTATGAAAA ATAGTACAGA GGTTCCTCAG
64451 ATATTAAAAA TGGAATTATT GTACTATCTT GTTGGAGGTC AAAAGAATGA
64501 GTGTTGTGAC CAACTCATTA TACCACTGGA GGCTATATGA GCAAACAGCA
64551 AACTGTTCTC ATGAATGCAG GATGTTGGCA AGCTGACAAC TGCATCTGCA
64601 ACCAGAAGGA ATGCTGAGGG CAGTCATGCC CCAGGCACAG TGTTTCTTGT
64651 GGTTATCTAT AGGAACATCT GGAGCCTGTT GTACAAAGAA ACCAATTATG
64701 TGAGCCTGTG ATAAATCAGG CAGCTGACTA ACCATTACCT GCTTCCTGCC
64751 CTGTTGATTC TACCTAATGA ATACAAAGGG CTGTATAAGC TCAGGGCCCT
64801 TGTTCCCTAT AAGCAAGGAG CCCCCTGACC CCTTCTTTAA AACAGATCTT
64851 TTTGTCTTTG TCTTCATTTC TGCGTTTGTC CTTCTTCTTC AGTCCTGAAC
64901 TGACAGCCAC AAGTGGCACC TGAACAGGGA CTTGAACAAA GAAGGTCTGC
64951 TGGAGCAGAA AAAGTGAAAC TGACCAGATG AATGAGAAAC CCTGGGATGA
65001 GTCTGCCTGC AGAGGATATA AGGTCAGTGT CCTAAAGAGG TACTGGGAGT
65051 GGGAAGTTTC TGAATCAGGG TAACGTGGGG GCAGAGTTTG TCTGTTGAGG
65101 AGCAGCATTA CGTGCAGTTG CTTAAAGTTT TACGTAAACA ATCTGGTGCT
65151 CAGGTTAGTT CTCAAACGCT GACTAAGCTG CTGCAGGAGG TTATCATGCA
65201 TAACCCCTCG TTTCCGCAGA CAGGCGCTCT TGATGTGGAA AATTGGAACT
65251 GAGTAGGGGA AGGATTAAAA TGGGCTCATC AAAAAGGTCT TAAAGTTTAT
65301 CCTTCCTTTT TTTTCTGTTT GGAGTTTAGT CCATACTGTC CTCCTGCCAT
65351 TATCTCATTC TTATTCTGCC AAACCGCAGG AGCCATGTTC TGAATCTCAA
65401 ATTTTGAAAG AATCTTTTGT CCCACCCACA ACACCCAAAG AAAATAATAA
65451 ACAGGAGAGG GAGGATGAAA ATTGGCGTCT ACCACCCCCT CCAGTAGCAG
65501 AAACACCTGT ACCATCTCCT TCAGTAACAG AAATAGAGAC CCCACTGCAA
65551 AGAATTCCGC GGACTGCTAC CATAGCTGGA GAGCCCTTAG GACATTGCAC
65601 TTTCACTATT TCTGTAAGGC CTGATCCAAA TAATCCACAG CAGTTTATTT
65651 ATGAACATGC CTCACTAGAG TTTAAGTTGT TGAAGGAATT AAAAGCTAGT
65701 GTAGTGAATA ATGGAGTACA GAGCCCATTT ACTTTAGGAT TGTTAGAATC
65751 TGTATTTTGA ACTATGTGTC TTCCATCCTT TGATGTAAAG CATTTGGCTC
65801 ACACTTGTTT GTCTGCTAGT GCATATCTGA AATGGAATTT AAATTGGCAA
65851 GAACTGTGTG CACACCAGGC TAGCACAGAA TTGTGCTGCC GGGCACAGGG
65901 GACATTACAG AGGATATGCT GTTGGGTAAT GGCCCTTATT CAGACCTGGA
65951 ATATCAAATG ACACTCCCAG ACGCTGCTTA TAAGCAGCGT GCACTGGCTG
66001 CTAAATGCAC CTGGGCCACA ATTCCAGAGG AAGGGGTCCC AATACAATCC
66051 TTTTTACATG GCATGCAAGG GTCAAAGGAG CACTATGCAC ATTTTCTTGC
66101 ATGATTACAA GAGGCAGTGA GGCATCAGAT TCCTCATACC ACTGCTGCAG

66151 AAATGCTAAC CTTAACTTTA GCTTTTGAGA ATGCAAACAC GGATTATAAA
66201 TGTGCACTGG CTCCTGTGAG ATGTACTAAA AACTTAGGAC ATTTTCTCAA
66251 AACTTGTCAA GATGTGAGAA CTGAGCTTCA TTGCTCTACA ATGTTAGCTC
66301 AAGCAATGGC TAATTTAGTA GTTAACAAAT CTAAAAAGGG CTAAGGGTCA
66351 AACCCTAAAA TGGGAAAATC TTATAATTGT GGAAAAATCG GACATTTCAA
66401 AAAGGAATGC CATCAGACCT TGGGCAAAAG GGATCTTATA ATGCAATACC
66451 CAACCTTCAG CAGAAAAAAT TCCAGAAATT TGCCCTTGTT GCAATAAAGG
66501 AAATCATTGG ACTAATCAAT GCAGTTCAAA ATTTCATCAG AATGGCACCC
66551 CTCTGTCGGG AACCAAGAAG GGAGCCTGGA CCCGGTCCCC TCAAACAATG
66601 AGGGCATTTC CTGTCCAGGC CACAACCCCA TTTCAGGGAG GAGTCTATGG
66651 AGGAACATTG ATTCCCTTTC CCCAGGAACA CCCAGAAGCA CAGGAAATAG
66701 ATCTCCCTGT CAGAGAATGG GTTACATTAG TTGGAGGAAA CAAACCCACT
66751 AAAATTCCCA CTGGTATTTG GAGACCTTTG CCAACAGGAT ATATGGGATT
66801 AATTTTGGGT AAAATCCATC TCAACTTACA GGACATTACT GTAGTCCCAG
66851 GAGTTGTTGA CTGTGATTAT GAAGGAGAAA TTCAAGTAGT GGTAATATCA
66901 CAAGATTTGT TAGTTTTTGA ACCTGGAGAA TATGTAACTC AACTACTGCT
66951 TATTCCCTGG GAGTTGTTTC CTTCTCCACA TAAGGAGAAA TGAGAGAATC
67001 AAGGATTTGG GAGTACAGCT AGGAGGAAAA TTTATTTATC ACAACCCATA
67051 GCATCTAATA GACCCACCTG TACAGTGCAA ATTAACGGAA AAAAATTTCT
67101 ATGGGCTTAT GGATATGGGA GCTGATGTGT CAGTAATATC TAAAAACAAT
67151 TGGCCCCCAT CCTGGCCCCT GCAATTAACT CCTACATCGC TAGTGGGAAT
67201 AGGAACAGCT CAAAGTGTTC AACAGAGTGC TGAAATTTTA CCCTGTCTCA
67251 AACCAGATGG ACAGTCATGT ACTTTTAAAA TTTATTTTGC AAATGTAACT
67301 GTTAACCTAT GGGGCCAAGA TTTACTTACA GCATGGGATA TAAGACTTGC
67351 AAATGAAACT ATTGACAATC CAGGGTTCAA AATGTTAAAG AAAATGGGAT
67401 GTCAGGCAGA AAAGGCTTAG AAAAGTCCCT ACAGGGAAAC ACTGATCCTA
67451 TATCAATAGC TGGGCAAACA GATAGAAAAG GGCTAGGTCA TCAGAATTTC
67501 TGATGGGAGT CACTGATATT TCTCCCCCAT CTACTGTTTT ACCGCTGGAG
67551 TGGCTGACTA AAAAACCTGT ATGGGTGGAT CAGTGGCCCC TATCACAGGA
67601 GAAACTAACA CAATTCCATC AGCTAGTAAA AGAGCAAATG GATGCAGGAC
67651 ATATTGAAGA GTCAGTTAGC ACCTGGAATT CATCAGTATT TGTAATTCCT
67701 AAAAAGTCAG GAAAATGATG ACTGCTACAT GATTTGAGAG CTATTAATGC
67751 ACACATTAAA CCAATGGGTG CATTACAGCA AGGTCTGCCA TCCCCGGCAG
67801 CCTTTCCAGG AGGCTGGCCT CTCAAAGTAA TATATCTTAA AGATTTTTTA
67851 TTTATTTTTT TATTTTACTG TTACATGAGC AGGATAAGCA TCGATTTGCC
67901 TTTTATGTGC TTTCTGTTAA TCAAAAAGAG CCTGTCTCTC ATTATCAATG
67951 GAAAGTCTTA CCCCAAGGCA TGCTTAACAG CATTATATCA GCATGTTGTA
68001 GGATAGGCAT TAAAGGTGCC TCTGAATATG TTTCCCACAG CCTACATCCG
68051 TCATTATATG GATGATATTC TTTCTGCCCC TCCTACAGAT CAAATTTTAT
68101 ATCAGTTATT CAGATAAATA AAATGAGCTT TGACTTAAAT GGAATCTCAA
68151 AATAGCTCCA GAAAAGGTGC AAACAACCTC CTGATACCAG TACTTAGGCA
68201 CTATTGTTAC TGAAAGAAGT GTTTGGCCTC AGAAAGTAGT CCTCCATAGG
68251 GACAGATTAC AAACTTTGAA TGATTTCCAA AAATTATTAG GGGACATTAA
68301 CTGGCTGTGC CCAATGCTAG GTATTCCTGC TTATCAACTC AAACACCTTT
68351 ATCAGACCCT TCAAGGAGAT TCTCCATTAG ACTCTCCTCA GCAACTTACT
68401 AAGGAGGCAA AAGCTAAGTT ACAACTTGTA GAGCTGATGT TTTGGCAACG
68451 ACATGCCTCC TGGCTACAGC CACAAAAGGC TTTGCTTCTG TTTATTCTTC
68501 CTACCCCCCA TTCAGCAACA AGACTTTTAG GCCAATTCAT AGTCAAATCT
68551 GTAGTAGTAT TAGAATGGCT TTTTTTAAAT CCAATCAGAC AGTGAAATCT
68601 TTGCAAGTTT ATCTTTCTTT AATTACTCAA CTTATAACAA TAGGTAGGCA
68651 TAGATCAAAA ATGCTTATGG GATATGATCC AGACAAAATT ATTGTTCCCT
68701 TGGATTCCCA ACAACACACT GCAGCATGGG AAATGTTGAC TGCATGGCAA
68751 ATTGCTCTTG CAGATTTCAT AGGAATAATA GATAACCATT ATCCATCAGA
68801 CAAAATTTTG CAATTTTATA AAGTTCACCC TTTTATTCTC CCTGTAATCA
68851 CTCATCACAA GCCTATTCCA GGTGAACAGA CCTATTTTAC TGATGGTTCT
68901 GCCAAAGGAC ACGCAGCTAT TTATGGACCT AACATACTTA GACAATAAAG
68951 ACCTCTGGAG CTTCAGCTCA ATGCTCAGAA TTAATGATAG TTATTCAGGT
69001 TTTACAGCTC ACCACTTCAT CTCCTAATAA CATTGTTTGT GATTCAGCCT
69051 ATGTTGTAAA TGTAGCCAGT CGTGCTGAAA CTGCCACTAT TAAGAGCACC
69101 CTAGAACCAG AGCTGCTTAA CTTGTTTCTA AGACTTCAAC AAGCTGTTCG
69151 CTCTCATGCT ACTCCTTTTC ATATTTCTCA TATTCACTCT CACACGCAAC
69201 TTCCTGGACC ACTATCTCTA GGTAATGATA AAGCAGATAA ACTAATCGGT
69251 TCTGTATTTC AACAAGCCCA AGCTTCTCAT GCATTACTGC ATCAAAACAC

```
69301  CTCTGCCCTT ACTCGTATGT TTCATCTGCC TCATGGACAG GCTGCAGCTA
69351  TTGTGCAAAC CTGCCCCACT TGCCAGCATG TTCCTGGTGT TGCACTTGTG
69401  GAAGGATGTA ACCCACGAGG CTTGGCACCA AATGAAATCT GGCAGATGGA
69451  TGTTACACAT ATAGCAGCCT TTGGGAAACT CAGCTGTGTT CGTGTGACTA
69501  TAGACACTCC CATATGCTAC ATGTCACATG CCAAACAGGA AACAGCTGGC
69551  CATGTCCAAC AACATTGTTT GTCATCATTC GCCCATATGG GGGTCCCTAA
69601  ACAATTAAAA ACTGACAATG GACCTGCTTG TGTTAGTCAT GCTTTTCAAA
69651  ATTTTTTACA GTTGTGGGCA ATCACTCATA ACACAGGAAT TTCTTACAAT
69701  TCTCGAGGAC AAGGCATTAT AGAGTGGGCA CATCAAACAC TACAGTGTAT
69751  GTTGTAAAAA CAAAAAGGGG GAATAGGAGA CAAGCTACCA CCTCAAACAA
69801  AATTACATTT ATCCTTATTT ACTTTTAATT TTTTACTTTT GATATGGATA
69851  GTAAGACTCT GGCCAAACAA CATTGGCAAA TGTTAGAGGG AAAGAGGAAA
69901  GTTTACCCAA AGGTACTATG GAAATCCCCA GAAGAAGGAC AATGGAAAGG
69951  CCTGGTGGAT TTACTGACGT GGGGATGAGG GTATGCTTGT GTTTTTACAG
70001  GAGATGGATA AACCGTGTGA GTGCCCTCAA GTTGTGTGCG ACCATGGAAT
70051  GGGAGACTGG AGGGATACAT GGATCCCAAC TACAGGCCCA GCTCCTCCAG
70101  TATGAGCCAT GAGCCAGTTG AATCTGAATG TGAAGATGGA ATGAAGACCG
70151  ACGAGAGTCA CACTGACGTC AACCCTCATA ACATGGGGTC AGATCAAGAA
70201  AACCACACCA GAAGCTGAGA AACTGGTGTA GTGCCAGGGT CAGGCAAAAA
70251  CCCCTGACTC CATGTTTATG GCCATGCTAG CTGTAATATC CTGTGCAGTA
70301  TGATTTTTCT GTGCAGAAGC AAAAACATAT TGGGCATATT TTCCTAACCC
70351  ACCGGTAGTG TGATCATACT CTGAAGCAGC ACTCCTCCTG AGATATATCA
70401  TGATCAAGGA GCATCAGTAC CAGGACCTCT AACTCCCCCT GACACAGAGC
70451  AATTAGACTC TCATAACAAT GGTATCAATT ATACCACTCC ATTGGAGGGA
70501  CTTCCTTTAT GTGTCACCCA GGATACATTG CTCAACTGCA GTTGCCTTGC
70551  AGTTTGATCC CAAGCATGGT TGAGTTACCA TAAAAAAATT ATGTACCTAT
70601  TAGACCTTAG CTTTATTAAT ATTACTTGTG TAGTTACTAA TCACTCCTGG
70651  CCCCATCACC CAAATTGTAC TGATTATACA GAATGGGCTC CCTTTGATAA
70701  TTCTCACCCC CCTCCTTGGG CCCACTGTCT TGGCCCCTTA GCTAGACAAT
70751  AGTCCATGTT AATGGGAGAC ATTATTGACT GGGGTCCCTG TGGTCATTAA
70801  GATGGGAGAG ATGAGAATCA GACCACATGG CATAAACTTC ACTGGCACTG
70851  GTGGCGAAAC TTTAACATCT CTTCACTTCA ACACACTGGG ATTCAATCCC
70901  AATCTGCCAT GCAACTTGCT TGGCATGGAA CGGGCTTTAG CCCACCTTTG
70951  CCTCAATGGC ATTATCAAGG AAAGAGAGGT CCAATTCAGG AGTCTATGTG
71001  GAAGGCAGCA CTCCCATATA TGAATGGCAG CATTTGGGTT GGGACACTAT
71051  CCAATAATAG TAATAGTGCT CAATACAGTT TAATGTTACC TTTGTAAAAA
71101  ATGTTTGAAA TTTGTGTTTT TAATCCCTAT GTTTTTCTAG CAGCAAAAAA
71151  GGACCAACTC CAGGTAAACA ATGCCCAATT GAATTGTGAT TCCTGTCAAC
71201  TCTATCATTG CCTTAATCAT AGCACAATAC AAACACACAG CATATCCACC
71251  CTAATAATTC TAGGTCGCAT TCCTGGATTA TGGATTCCTG TAAATCTATC
71301  TGAGCCTTGG GCAGCCACCC CCACTTTACA TTTTGTAAAA CTTCTTACTC
71351  AGCTTACTCA TGGCACTCGT AGAGCCTTAG GCATGATAAT TTTTACTATA
71401  GTCTCCTTAA TTACATTAAT ACCCTCTGTT GTGGTGTCCT CAGTAGCACT
71451  GGACAGCTCC ACTCAAACAG CTCAATATGC AGAAAATTGG ATGCATACAG
71501  CTGACCAGGC ATGGATGTTT CAAAATAAAA CTAACACTGA GATACAAACA
71551  GAAGTGGCAA TGTTAAAGAC TACTGTTCTG TGGCTAGAAG AACAAGTACA
71601  AAGCTTGCAG TTGCAGTAGC AATTGCGTTG TCATTTTAAC CATACTCATA
71651  TTTGTGTAAC CAATTAGGAA TATAATCAAA GTGAATATCC ATGGAACCTT
71701  GTAAAGGCCC ATTTACAGGG AGCTGTTACA TCCAATGTTA CTTTTGATAT
71751  TAATGATTTA CAAAGTAAAA TTCTAACAGC ACCTCAATAT CTTTTTCATA
71801  ATTATTGGAA TAATGTTACT ATGTTTCTGT TTTTTGTTCA TAGTCTGTAA
71851  AATCAACTGG AACACCAACC AGCAATTGAG AGCTGAACAG CCTGCAATTA
71901  CCTTTATTCA ATTAAATCAA AAGCAGAAAG GGGGAGATGT TGGAGGCTGA
71951  AAGAATGAGG GTCATGACCA ACTCAGTATA CCACTGGAGG CTATGTGAGC
72001  AAACAGCAAA CTGTTCTCAT GAATACAGGA TATTGGCAAG CTGACAGCTG
72051  CATCTGCCAC CAGAAGGAAT GCTGAGGACA GTCATGCATC AGGCACAGTG
72101  TTCCTTGTAG TTATCTATAG GAACATCTGG ACCCTGTTGT ATAAAGAAAG
72151  CAATTATTTG AGCCTGTGAT AAATCAAGCA GCTGACTAAA ACTGTTACCT
72201  CTTCCTCCCT GTTGATTCTA CCTAATACAT GTGAAGGGCT GTATAAGCTC
72251  AGGGCCCTTG TTCCCTAGAA GCAAGGAGCC CCCTGACCCC TTCTTTACAA
72301  CAAATCTTTT TGTTTTTGTC TTCATTTCTG CATTCATCCT CCTTCGTTCA
72351  GTCCCGAACC GACAGCCACA TGATCTGGCA ATCCCATTTC TGGATATCTA
72401  TATAAATGTT CAAAGCAGGA CCTGAAAGAA ACATTTCACA CCCCTGTTTA
```

```
72451  TAAGAGATTT ATTCTAAAAA TCCAAAAGGT AGAAGCTACT TGAATGTCCC
72501  TTGACAGATA AATAAAATAA AATAAAATAT GATATATACA TATAATATGA
72551  TTTAAAAAGA AAATCTTGGG CTGGGTGTGG TGGCTCATGC CTGTAATTCT
72601  AGCACTTTGG GAGGCCGAGG TGGGCAGATC ACGAGGTCAG GAGATTGAGA
72651  CCATCCTGGC TAACACGGTG AAACCCCATC TCTACTAAAA ATACAAAAAA
72701  TTAGCCAGGC ATGGTGGCAG GTGCCTGTGG TCCCAGCTAC TCAGGAGGCT
72751  GAGGCAGGAG AATGATGTGA ACCCAGGAGG TGGAGCTTGC AGTAACCGGA
72801  GATTGCACAA CTGCACTCCA GCCTGGGCGA CAGAGTGAGA CTGTCTCAAA
72851  AAAAAATAAA TAAAATAAAT AAATAAATAA ATCACACACT GCAATGACAG
72901  TAAACCTTTA GGACATAATG TTAAGTGAAA TGTGCCAGGA AACAAAGTGA
72951  CAGTGAGTGT ATGATTCCTC TTATGATATA TCTTAAGTAG TCCAACTCAC
73001  AGAAACAGAA AGTAGAATGT CAAAGGCTCA GGAGAGGGTA AAATGGTCGG
73051  TTGACGTTTA TGGCTATTGA GTTTTAGTTT TGCAATGGAA AAGCTCTAGA
73101  AGCCTGTTGC ATAACAATGT GGATATATGT AACACTACTA AATTATGCAA
73151  TTACAAAGGT ATAGACTGGT AAATTTTGTT GTGCTTTATT ACAATTAAAA
73201  TATTGTAAAG TGATACATAA AAGAGATACA GAGTTATAAA CTTTTCAGAA
73251  AATTACCTTC AAATTATAAG CGTGTTTTTC TCACACAAAG ATAATATAGA
73301  TTCATCAAAA AATACATGGG CAAATTAAGA CTATTTACAT GACTACTCTC
73351  CTGAACAAGT TAAAACAAAC TTTTGACATC AGCCAAGAAG AGAAATATGC
73401  AAGATAAGAA TAAATGGAGT ATATTTATAG AGGCAAACAA ACACATGATT
73451  TTATTGGTGG TAGATATGAC TGATTCATAT TTTAATTAAA CCCCACATCG
73501  ACTCGATGTG TACATAGAGT TGCAGATTTA CATCCAAAAT CATAATATGT
73551  AGGTAAAACC AAATTCACAA AACACAAATG TCAAAGAAGC TACCCCAAAA
73601  AAGAAGCACA GTAATATGAA ATTTCAAAAC AAGAATGAGA GAAACATTAA
73651  CAACAACAAC AACAACAAAA ACACTTCTAT ATAAAACATA GATGTACAAT
73701  TAAGAAAGAA TCCGCTTAGA TAATTACAAT TTCCCGCTGT GACCTTGCAC
73751  TGGTGGTGAG CACAGATTTT GAATCATGAC TATGTTAGGG AGACGCCCAA
73801  GGAGACAGAC AGTGCCACCT CCAGAGAAGC CATTGCTTCT CCTCCTGCCG
73851  CTGCTGCTGC TGCCCCCACC GTCCGCTGCG CCTGCAGCCC CCACTGAGCG
73901  TCGGACTCCT TCCTGGAGTA GGGAGGTCCT GTTCCTTCTG GAGCGACAGA
73951  CACCCTTTCT CCTGGCCTTC TCGCTTACTA GCCCGGCAGG TGCTGGACAG
74001  GAGATCTGAG CTGGTCCTGC GTCTCTGAGG AGCTAGGAGC CCGGCTGGGA
74051  GAACAAGGAG ACGAACTGTG GGGAGAAGGG GCGACAGGAA CGCCAGGCTC
74101  ATGGGACCGC TGGCAGCGGC CTGGGTATGG CTGGCGGCTG AATGGTCAGA
74151  GATACGAGAG GTGGCCACTG TCCCCACCTT TGGCCCCCTA GCCGGCATTC
74201  GTACATTCTG TGCTCAACAA ACGGAAGCGG CAGCTGGAGC TGCTGCTCCG
74251  GGAGGTGGAG TGGCCTGGCA GAGGGCACAT GGCTGCCACC TGCTGCAAGG
74301  TGAGCTGGTC TGCAGCCTGG GCCCACAAAA GGCCGCTCTT GTGCAGGACA
74351  CACCGCTGCC CTTGACCCTC TTGCTCCCCC GCCTGCTGTG CAAAATGCTC
74401  AGGTCCCTGA TCTCGGGCTT TCCTGGCAAG TGCACTGTGG TGGGGAGGCA
74451  GCAGGGAGGA GGGCTTTTCC AGGAGCCCTG AACAGAGGAT CTTGGCATAA
74501  AGAGGAGAGA GAGGTGGCTG ACTGGTTCCA CTTGTAGGTA GGGGGGCAAC
74551  AAACCCCATG GGACCCTGTT TTTTCAGGGA GATTTCAGTT CACTTCTTAT
74601  CTTTTCTCCA CCCACTTGAG CCTCTGAGAA TAGAGGAGAC GAGGCTGTTT
74651  TAAATTGGCC TAACCATAAT GGTCTGGACC CTTGCCCCAG GGCAGACCTA
74701  ATTTTGGGGC TCTTTGCAGC ATGGAGGCTC ACGCCTGTCC ACCCCAGGTG
74751  TCTTCAATAT AGGGTCTAGT TAGGCCTGGC TGGCAGTGAT GCTGAGACGC
74801  AGCACGACCT GGCCAGATCT TCGCCTGTTA CAGGACATTA TAGCCTTCAG
74851  TGCCCTGTGC CATTTATCTC GCCTCCAGAA GCCCCTGTGA GCCTCAGTGT
74901  TGCCGGTGCC CAGGCCCTGG CTGCCTCTCT ATTAGGGTCC CATCTTATGC
74951  CTCCTAAATG CACCGGGGTC TCACTCTGCC TTTCTCCCTT TCCCAGAACA
75001  GGCCCCTTCA ACTCCAACAG AACATGCCTG GACCATGTGC ATCCCTCTTC
75051  AGTGTTAAAA CAAAGAAAAT TTATTTTTTT TCCACTGAAC ATGTAACTGA
75101  TTTAACATAT AAGGAGGTCA GCTTTATGCA TAGATCTATG CATGTAAATA
75151  TATACAAAAA TTCTAACGCT GTGGGAAAAT TAACATCCTT ACACTTTGTT
75201  CAGTTATTTT ATAGTTTTCT CTCTCTCACT CAATTGCTTT TTTTTTTTTT
75251  TTTTTTGAGA CAGAGTCTTG CACTGTTGCC CAGGCTAGAG TACAGTGGCA
75301  AAATCTCAGC ACACTGCAGC CTTTGCCTCC TGGATTCAGG AGATTCTCAT
75351  GCTTCAGCCA CCTGAGTAGC TGGAATTACA GGCATGGGTC ACCATGCCCA
75401  GCTGTTTCAT GTGTTTTTTT TTAGTCGAGA CCGGGTTTTG CCATATTGCC
75451  CAGGCTGGTC TCGAACTCCT GGTCTCAACT GATCTACCCT CCTTGGCCTT
75501  GCAAAATGCT GGGATTACTG GCATGAGCCA CCATGCCCAG CCTACCTGTC
75551  ACTATCTCTA TGTTTATTTG TTCAACAGGA AAATTCTCAG TGAAGACTCC
```

75601 TCAGTATGAA GGAGATAAGC CTGCACAATC AGTCACTGAT AGATGCTTAG
75651 TGGAAAAACT TCCAATTCCC ATTTACAGCT CTCAGAGCTA GGATTAAAAA
75701 CTCCTGGTCA TAAACTCATG TGATGAGAAG TTATAGCACG CCCTCATTTT
75751 CTACATATCC ACTTGCATTT ATGGTTGGCT TTTGAACTTG CTAGAAGGGA
75801 AAGAAGTGCA AATGTGTCCT CCTTAGAGCT ACTCTCCTCC CCTTGGTGGG
75851 TTTCCAGTTT GTGCATTGTC CAGATGGCCC AGGAGCTGAC GATCAAAGGG
75901 AAGAAGTCAT GTTTGTCATG AGAATGCTTT GCTGCATCAG GATTCAGTGA
75951 AGCTGTTCAC CGCCTGGAGC CCATGCAGCC TCAAGAGGCA GGATGGAGCT
76001 CAGAAACCAT CACTGAGGTT AGAAAGTGAG CACCAAAGTT GAGGGAAGCC
76051 CACAGGAGTG AGCCGAAGTG CTCCCTTTGG ATTTCCAAGT GGTTGCTGCT
76101 GCTTCTTCCA TCAGCCTTGC TTCTGACCAC AATGTGTTCC TGGTGCCTTC
76151 TTCTTGGCAT TTTGCTGTTT GTGTCCAAGG AAAATAGTCC TGCATGGCAG
76201 TGGTGAGAAG GATGGCTGCC TGCTGAAGCT GATTTGCTGG TAAGCTTTGC
76251 AGCCTGTTAA GCAGAGCCTG AAATTCTTTC TCACTGAGTG GTGATTCAAA
76301 CCTTGGAGGG TCCTCCCCTT GTGGATCGGC ATTCAGAAAA GATTGTGCCT
76351 TTTCCTGAAA CTCTGGGGAT TATAGGAGAT CATGCAGATG GCTGGTGTGG
76401 TTGCTCCAGC AGGTGGATGT CTCCTTCGTG GCTTGTGTCT GTTTCTGTCA
76451 CAGGGGAGAC TCAGTGTGCA TGGGCTGCTA AGGTGCTCCT GCTTCAGGTT
76501 GAATCATGAC ATCCTAGGAC CCCTTGCCCC GGCTCCACCA CTGCTGGGAT
76551 TTGGCCTGTT GACCCAAACT CTAAAATTGT GGCTAAGAAT GACCAGGCTG
76601 AGGCTGTCCT TTAAGGCAGA ACTTCCAGGT TAGTGTCTCA TTTTTCTTAT
76651 CCTGAAATTT TCCTTTCGAC AGAGGCCAGA AGCAAGTCTG TGTATGGGAG
76701 AGCCTCCCTC CTAGAGCTGG TACCATTGAC ACATGACTCC TGAGTGCCAG
76751 AAGAGGTTGA GAGAACTCTC CCATCTGCAC AGCCTGTCTC ATGCAGAATG
76801 CAGATGGATC CAAAAAATCA CAGGATGTGG GAGGTGAAGG AAGAGCTTGT
76851 AAAAATGAAA AGTGGCTGGT GAAAGAGTAG GAGGCATGAA GAGGATGAGA
76901 CCTGCCTGGG GCAGTGCACA TGTTTTGTTC CAGCCAAACA ATCAGATGAG
76951 GTCTTGGTCT TGGACCTGGT GCCAGGGAAT TCATAAGCCC CCTTTTGCTG
77001 TGGCCTGGGA GCTGAGGTCT TTGGTTCTGA AACCAAATGT AAATTTTGGA
77051 CTCTGGAATA CCTGTCTGTT TAACCAGTTT CTCTCTACTG GTGATCGCAG
77101 GAAATGAATT ATCCTGTAAA ATTTTGTGGA TTCTTCTCAA AGGCTTCAAT
77151 GAGTACACTG ATTTGTGGTT CAGTGATGGA TGTCCCTTTC CTCCTGCCTT
77201 CTTATTTGAC TTACACCATC AATATTAACT ATGGCAGTTA TGGTAATATC
77251 ATTCTTTACA ACAGAGGAAA CTTCAAGTCA TTCATTGATG ATATCAAAGC
77301 CTCATTCTCC TACATTAAAC ATTCTCCTAC ATTTAGCTTC TTGAATCTTC
77351 TATGTCCACT GTCTAGAAAA CCTAAACACA TGAAGTACCA CAAACTAGAA
77401 ATAAATACCT CAATAGGCAC CAATCATAAA AGTAAATCCA AGAAGGAACA
77451 GAATATATGA ATACATATAT AACAAGTAAA GGGATTCAAT TAATTAAAAA
77501 TCATCACACA CAAAAAAGCC CAGAGTCATG TGGCTTAACT GATAAATTCT
77551 ACTAAACATT TAGTGAAGAA TTAATGCCAA CTCTTCACAA GGCCTTCCAG
77601 AAAATAGAAG ACAGTTATTG GGAACACTTC CCAATTTGTT CTATCAGGCC
77651 AGTATTACCC TGATACTAAA GCCAGACAAA AGCATCACAA GTAAATATGA
77701 ACATAGATGA ATTTCCCTGA TAAATACACA AACAGAAAAT CTCAAAAAAG
77751 AGTGAAATGA ATCAAGAATA AATCAAAATG ACTGTACACC ATGACCAAAT
77801 GGAATTATTT CACAAATGCA ATATTGATCT ATCCAATAAT CAATCAATGC
77851 ATTACACAAA GTAATAGGAT AAAGGAAATT AACAGAAAGG TCCTTTCAAC
77901 AGACACAGGG AGCATTTAAC CAATCCAATA TTCATTCACG ATCTCCCTGG
77951 AAAAGAGGAG TACAAGAAAC TTCCTAGATC TGCTAAAAGG CGTCAATGTA
78001 AAACTTACAG CTAACCCCAT AATAATAAAA TACTGGTTGT TGTGTCTTGA
78051 CATTTGAGAA CAAGACAAAA ATGTTAACAT CCAAATAAAT TACATAAGAA
78101 AAATAAACAA AATCATCAAT GTGGGAAAAT AAGAGGTTAG AACTCTCTAT
78151 CATTGCAGAG GACATAATGT GAATATAAAA GTTTATAAGA AATTCATTAA
78201 AACTCACTAC AACCAATAAG TGAGTTCAGC AACATCACAA GATACAAAAC
78251 CAATATACAA AGTTCAATCG TACTTTTATG TACTAACAAT GATCAACCTG
78301 AAAATAAAAT TAAGAAAACA ATTCCATTTG TGTATGTATG AAAAGGAAAA
78351 AAAATATTTA GGAGTAGATT TAATCAATTG CAATTTTACA GTAAAAAGAA
78401 AAACTGTTAA ATAACTTTTA AAAACTGAAA AAAATAGGCA TCAAGGTTTA
78451 TTTAGTTAAT ATTATTCCAT TATTCAGTGT ATCCATTTAA ACTCCATACA
78501 ATAAAATAAA GTATATTTTA GCATTATTTT TAGTGTAATA AATACAGGTT
78551 GGTAGAAAAC TCATAAATAT AAAAGGAACT AATACTACAT GTGCAGAGTA
78601 GCCCATTCTA GTTTTTCTAT TTTACATCTA CTGTAAATCA AACAGATTTT
78651 CTTCCATATT TTATAGTGGA TCTTAATATA AAATCCAAAT TGTTAAATAT
78701 GTGAGATTGA TTATAAGCTT GCCAAGGAGG TTTTTACTCA GTGTGGGAAT

78751 TTGGAGAGCA TAAAGCTGCA AAGACAGAAG CAAATGTTTT TGAATGAATT
78801 ATGAGGGACA ATACTCACAG GAGTGATTCC CACTTTTATC AGTTGACTCA
78851 TGTGATCTTA CTTTAGGAGA AACTGACTCT CATTTTAGAC ATTGGTTCAT
78901 CACAGATGCT AAATGAACCA GCACCAAGTT TATATCCAGG AGAACTGCTC
78951 ACTGTAGAGG ATTTTGTCTC CTAGCTGATG ACATGTTCTG TCTTTCATAG
79001 GCTACTCCAA AACTTTTATA ATGATGGTAG AGTTCTGTAA AGTGGAGCCT
79051 CAGGCCTGCT ATCCCCATGC TCCTGGCCAG CAGCAGCATC TCCCCTGAGC
79101 ATGGTGACAT AGGGCATGCC ATTGACACCC AATTTGCTGG TGTTCACCTC
79151 CACATACTAA GTTGCTAGAG AATTTGCAGT GGATTTCCAT TTTGCTGCAT
79201 CTGGCTGTCC AGAAGCCCTG CAAGTAGAAT GGGATGGTCA GGAGAAAACA
79251 TTGAAAGAAT AGATGGGAGT TCAGAGGCTG CCCACCCTCC TGCCCTCTGC
79301 CCACGGGCCA CAGCCCTCAC CCAGCTGTCC AGTGTGTATG TCTGCTAAAG
79351 GCTGCTGCAC TTGTTCTCCA TCACAGAGGT GGGGTGACAG CAGTCTGTGG
79401 GCACCACACT CCATGATCAG CCTCTACTGT TGGTGCCACA GCTCCAGGTA
79451 GAAGGGCAGG TGAGCCAAAG ACAGGTCCCA CCTTCCACAT CCAGCCCACA
79501 TTCCCACCAA CTTCCAGGCC CACCTCCATA TGCTGATGTG ATGCTCTCCT
79551 TAGAGCTCTT GTGGTTCTTC AGCCAGGAGA TGGAGAGAGT GGGGTTTCAA
79601 GAGTAAGGCA GTGAAAGTTG GTTTGGGTGG CCACCATGGT TGGTGGTTTC
79651 CTGTCCATCC ACTCGGTCCA AGTCCAGTAA GGGGCCTCTG CTGGTGGAAG
79701 CACACATGAA GGCCATAGCT GGGGTGAGGA GCAGAGACTT ATCTCACCAG
79751 ACCCCCAATT CAAGGTTGCC GCTGCCCACC CACACCCTCT CTTTTCCCTC
79801 CTGTGTAGGG AGATTGTTGG CCTTTCAAAC CCCTCTTCCT GGGCTGAATA
79851 AGGATTCCAG GAGCTTCAAC ATAATGTCCC CACCCAGTCA TGCTCAGAGC
79901 TGGGCCATGT GTCCCCTCCC ATTCCCTTCA CTTCCCACAA GTGGCTGCTC
79951 CTGCTGAGAG GTTGGGGTGC TTCATCCTGG CCTAAAAACC TCAAAGAATA
80001 ATGGAGTCTC CAAAGGAGCC CCCACCCACC CAGGGAGGCT GACAGGGAGG
80051 GTCCACCAGG AAGGGAGCCC AGCAGGTAGC CCAGCTAAGT GAATGAGCCA
80101 GGGTAGGCAT TGGGAGCAGT TTACCAGGAG AAGAAACTCA GCCCCTTGCA
80151 GAGCGGGGAG CCTCAGAAGC AGCAGAGAAG CCTTGCCCCA CAAGACTCTG
80201 AGTCCCTAAG CCACCCCCTG TAGAGCTCCA GGGCACTGGT GAGGATGGCC
80251 TCCTGGAGGC CTCAGCTCTT TTTTGTGCTA ATGTCCAGGG CTGTCACCAT
80301 TCCGCCCCCG CCCCCCTCAC AGCTGAACTA TTTTTCTTCT TTCTGGGGCT
80351 GGGGTGGGGC TGCCTTCCTG CCTGGATTCA TGGGTAGGCT GGATTGCCTT
80401 ACCCCCAGGA GAGAGGCACC AGGGGCCCAG AAAAGAGGAA GGAGGCAGCC
80451 CCTTCCCCAC AGTGACCTCC TCGCAATTCA CATGCAGCAA CAGGCCCACC
80501 TCTCAGAGTA AGATGCTGAG GCACAGGTAG GAGCTGTGTA GAGAGACACT
80551 GGGAAGAGGC ACCTCTCACA GCTCTGAGCT GACCTCCAGC CCTCCAGGGA
80601 TGGGGGAAGG TAGACCCATT GGTGATGAAG ACAGCTCAGT GAGCTGTGGC
80651 AGAGGAAGTT CCCAGGACTG GGAGACGACA GTGACTCAAC GCTGCTCATT
80701 TCTAGACTGT GCTTTCTGAA AGTGGCCCTT CAGTTACCCC CACAGCTTGA
80751 GGCCACACAA GCCTGGGATG GCCAGTTAGG CAGACGCAAG CAGGGATTCA
80801 GGGGGAGTGC ACTAGGGTGT GTGGGCAGGG GCAGAGGCCA TTGAGGCAGG
80851 TGAGGAGAAA TTTTCATCCT CTTCCTGGTC TGCCCCTCTC CTGGGGTCTA
80901 ATTTCCTCTA TTCCGTCTGC CTCTGGCTCC CTGGCTGGCT CCTCTTCACT
80951 CTCTTGCCTG CTCACCCCAG AGGTCCCAGG GGCTCAGCCC ACCACAAATG
81001 GTCCCCAAGT TGTAGCTGAC CCTTCCATGT CCATCCCATG AGGACCCTCA
81051 TCTGCCTGAG TATATCTCTG GGCTCCTCTG AAACCAGAAG TCCCACCTCA
81101 CTGACTGCTC CATGGCTAGG CAGCATCCAC CTGCCACTGT TCCAGGCCAG
81151 AATGACTGGG CATTGTCCCC CTGCCTGCTC CCTCACACCT ACAGCTCATG
81201 CCCCCAAAAT GCTGTTGGCA TCCTTCATAC AACCCTCACA GCCACCCTGC
81251 CCCCTGCTGG GCTGTAGGTT TGGTCTCCTG GTGCCTTAAC CCCTTTGTCC
81301 ATCTGCCCCT GGGCAGCCAC CTGAGCCCCT GAGCACTGCT GTGCTCACAT
81351 GTGCAGTAGC CCCCTCACCC AGAGCCAGCA TCGAAGTCTC CACAGGCCAA
81401 TTCTGGCCTC ATCACTGCTC CTGGAACCCC AGGGCCCTGT GCCCCAATCT
81451 CCCCATCTGC AGCATGGGTG TCTCTTCCTA CCCCAAGCCT GCCCCCCAGA
81501 GCTCAAGACA TCCAGAGCCA TCTAATACAT ATGTAATACA TACAAATTAC
81551 ATAACAATTT GTAATATGTT GTACTACATA CACATTTTCA TATGAATTCA
81601 TCATAATACG TACAAATTAT GATGTCATAA TATATTGTGA TGTGACAATA
81651 CACATGAATT ATCATGTCAT AATACATTGT GATGTCATAA CACATACTAA
81701 TTATGATGTC ATGATATATT GTGATGTTAT AATGCATATG AATTATGGTG
81751 TCACAATACA TATGAATTAT GATGTCATGA TACATTGTGA CGTAATAAGA
81801 ATTGTGACAT CATAATATAT CATGATGTCA TATGCATGCA ACTTATGATG
81851 TCATGATATA CTGCAATGCC TTAATACAAA CCAATTATGA TACAGTAATA

```
81901  TGTTGTGATG TCATAATATC ATATTTATTT ATCATATTTA TCATATAACA
81951  TTTTGTCAGA TATTTTTATA AGAAAATTGA GTGAAATTTT GTAACATTAA
82001  CATATAAAGA AGCTTACAAT CATGATGAAA GATGAAAGAG GAGCAGGCAT
82051  CTCACGTGGT AGGAGTGGGA ACAGGAAAGA TGGGGGAGAG ATACGCCTCA
82101  CTTTTAAACC ACCAGATCTT GTGTGTACTC ACTGTGACAA TGACAGCACA
82151  GAGCCATGAG AAATCCATCT CTATAATTCA TCCACCTCTC ACCAGGCCCC
82201  ACCTGTAACA TCAGGGATTA AAATTCAATA TGAGATTTGG AGGGGAAATC
82251  TAAACTATAT CATATGACGA TTAGAAAAAC AGATGAGGTC CTTCACGTCT
82301  CTTTGAAGCA ATTGTGAATG GGAGTTCACT CATGATTTGG CTCTCTGTCT
82351  GTTATTGGTG CATAAGAATG CTTGTGATTT TTGTACATTG ATTTTGTATC
82401  CTGAGACTTT GCTGAAGTTG CTTATCAGCT TAAGGAGATT TTGGGCTGAG
82451  ACAGTGGGGT TTTCTAGATA TACAATCATG TCATCTGCAA ACAGGGACAA
82501  TTTTACTTCC TCTTTTCCTA ATTGAATACC CTTTATTTCC TTCTCCTGCC
82551  TAATTGCCCT GGCCAGAACT TCCAACACTA GGTTGAATAG GAGTGGTGAG
82601  AGAGGGCATC CCTGTCTTGT GCCCGTTTTC AAAGGGAATG CTTCCAGTTT
82651  TTGCCCATTC AGTATGATAT TGGCTGTGGG TTTGTCATAG ATAGCTCTTA
82701  TTATTTTGAG ATACGTCCCA TCAATACCTA ATTTATTGAG AGTTTTTAGC
82751  ATGAAGGGTT GTTGAATTTT GTCAAAGGCC TTTTCTGCAT CTATTGAGAT
82801  AATGATGTGG TTTTTGTCTT TGGTTCTGTT TATATGCTGG ATTACATTTA
82851  TTGATTTGCG TATATTGAAC CAGCCTTGCA TCCCAGGGAT GAAGCCCACT
82901  TGATCATGGT GGATAAGCTT CTTGATGTGC TGCTGGATTC GGTTTGCCAG
82951  TATTTTATTG AGGATTTTTG CATNNNNNNN NNNNNNNNNN NNNNNNNNNN
83001  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
83051  NNNNNNNNNN NNNNNNNNNN NNNTTTTAGG GCAGGCCTGG TGGTGACAAA
83101  ATCTCTCAGA ATTTGCTTGT CTGTAAAGTA TTTTATTTCT CCTTCACTTA
83151  TGAAGCTTAG TTTGGCTGGA TATGAAATTC TGGGTTGAAA ATTCTTTTCT
83201  TTAATAATGT TGAATATTGG CCCCCACTCT CTTCTGGCTT GTAGAGTTTC
83251  TGCCAAGAGA TCCGCTGTTA GTGTGATGGG CTTCCCTTTG TGGGTAACCC
83301  GACCTTTCTC TCTGGCTGCC CTTAACATTT TTTCCTTCAT TTCAACTTTG
83351  GTGAATCTGA CAATTATGTG TCTTGGAGTT GCTCTTCTCG AGGAGTATCT
83401  TTGTGGCGTT CTCTGTATTT CCTGAATCTG AATGTTGGCC TGCCTTGCTA
83451  GATTGGGGAA GTTCTCCTGG ATAATATCCT GCAGAGTGTT TTCCAACTTG
83501  GTTCCATTCT CCCCGTCACT TTCAGGTACA CCAATCAGAA GTAGATTTGG
83551  TCTTTTCACA TAGTCCCATA TTTCTTGGAG GCTTTGTTTG TTTCTTTTTA
83601  TTCTTTTTTC TCTAAACTTC CCTTCTCGCT TCATTTCATT CATTTGATCT
83651  TCCATCACTG ATGCCCTTTC TTCCAGTTGG ATCGTATCAG CTCCTGAGGC
83701  TTGTGCTTTC TTCACGTAGT TCTCGAGCCT TGGCTTTCAG CTCCATCAGC
83751  TCCTTTAAGC ACTTCTCTGT AGTGGTTATT CTAGTTATAC ATTCCTCTAA
83801  GTTTTTTTCA AAGTTTTCAA CTTCTTTGTC TTTGGTTTGA ATTTCCTCCT
83851  GTAGCTCGGA GTAAGTAGTT TGATCATCTG AAGCCTTCTT CTCTCAACTC
83901  ATGTGGAGAA ATAGGAACAC TTTTACACTG TTGGTGGGAC TGTAAACTAG
83951  TTCAACCATT GTGGAAGTCA GTGTGGCGAT TCCTCAGGGA TCTAGAACTA
84001  GAAATACCAT TTGACCCAGC CATCCCATTA CTGGGTATAT ACCCAAAGGA
84051  CTATAAATCA TGCTGCTATA AAGACACATG CACACGTATG TTTATTGCGG
84101  CATTATTCAC AATAGCAAAG ACTTGGAACC AACCCAAATG TCCAACAATG
84151  ATAGACTGGA TTAAGAAAAT GTGGCACATA TACACCATGG AATACTATGC
84201  AGCCATAAAA AATGATGAGT TCATGTCCTT TGTAGGGACA TGGATGAAAT
84251  TGGAAATCAT CATTCTCAGT AAACTATCGC AAGAACAAAA AACCAAACAC
84301  CGCATGTTCT CACTCATAGG TGGGAATTGA ACAATGAGAA CACATGGACA
84351  CAGGAAGGGG AACATCACAC TCTGGGGACT GTTGTGGGGT GGGGGGAGGA
84401  GGGAGGGATA GCATTGGGAG ATATACCTAA TGCTAAATGA TGAGTTAATG
84451  GGTGCAGCAC ACCAGCATGG CACATGTATA CATATGTAAC TAACCTGCAC
84501  ATTGTGCACA TGTACCCTAA AACTTAAAGT ATAATAATAA TAAAAAAAAG
84551  AAAGAAAAAC AGATGAGGCT GGGTGCAGTG GCTCACGCCT GTAAACTCAG
84601  CACCTTAGTA ATCCGAGGTG GGCAGATCAC AAGGTCAGGA GATTGAAGCC
84651  ATCCAGTCTA ACACGGTGAA ACCACCCCCA TCTCTACTAA AAATACAAAA
84701  AAATTAACTG GGTGTGGTGG CACTCAGCAG TAGTCCCAGC TACTTGGGAG
84751  GCTGAGGCAG GAGAATCGCT TAAACCCTAG AGACAGGTAT TGCAGTGAGC
84801  TGAGGTCATA CCACTGCACT CCAGCCTGGG CAAACAGAGT GAGAATCCAT
84851  CTCAAAAAAA AAAAAAAAAA AAGAAAAACA AAAAAAGAAA AGAAAAACAG
84901  ATGAAATAGT AATGGTTTCT ACAGATACTT TCATTACAAA GCATATGGAT
84951  AAATGATTTA ATTCAACATT CTGTGGTCAG AAGAGAGAAG GGAGGTGTAA
85001  AGGGGACTTT GACTGCATTT GTTATACTTC CCTTATGCTG TTGTTATGAG
```

```
85051  TTTTGATGTC ACCACCTGAA GGGGTATTCA TGGACGGAAG AATTATTGCT
85101  ATTGTTGTGA TTATTCCTTT TTCTTTTGTA TTAGTAAAAT AAATCTTTTA
85151  GCTTCTCATA TAATTTTCTT TAAAAGCCCT AAGAGTTTCG GTTAAATTCT
85201  TCGTTATTGT GTGTTATAAA AATTGACAGG GAAATGGCTA AAATAGGTTA
85251  AAATTACACA AACTCTAGGA GTCCAGTTTC TGTTAGACAG GCTTAGGAGA
85301  GACAGAACTG GAAACACTCC ACCATCATAG ACATCAGACA CATGGGGCTC
85351  ACTTTCTGTC CCAGCCCTGC CCAGATCCAC CCTCTTCTAA GGCCTTATCC
85401  AGGCCTGGCC TCACCCTAGA ATCTCCTCTC ACAGAAGGAG ATCAAAATTA
85451  AAGGAGATCA AAAATTTGAG TGGTGGCTCC TCCTGCCTCT CCTTAGCTGA
85501  TGCCCACAAT TTCCTGAAAA TAAAAAGCAG ATAAATGGGA GCAAATGATT
85551  ATCTATTTGT GGGTCACAAT TTCTTTTTCA TTGAAGCCAG TGCTTGTAGA
85601  GACATCCCAC CTAGCAAACT GTTTTTCTGC CGATCCGGTA GATGCTCTAC
85651  AAAGCACAAG AAAGTCAATA TAAATACCAA AAATCCCTCT GAACAGTTCA
85701  CCCTTTCTGT ATCCCTTCCA TCTGTCTATA TAGATTTTAT TCTACACATT
85751  TTCTTTTTAA GGTAAGTAAT TTTAAAAATG AAAGAAAAAT AGAAATGCTG
85801  GGCCCTTCAT TTAAAGCCTA GGAATTACAG AACACTTAGC AGCCAACTAC
85851  CAGGGTTGAG GATTCACTCA CATCAGGTGA TGTTTGCAGC ACAGTGCTGT
85901  GTAACAGACT TCTGAACACA TAGTACACAC TCAGTAAACA TTGTATTAAC
85951  TCATGGATAC ATGTTTTTCA AATGCAGACT TACTCAACCA TTGATGCCTT
86001  CTCTAGGCTC TATAACCTTT AAAGAGCCAG CAGAGAATAA CATGTTTGTA
86051  TATAGTGATT GGTGGTTTCC ACTTTGGGCC AAAAGGGTTT GTGTTGTGGT
86101  AAGAGTGTTG GGTGTCAGGA ACTCTCTGTG CTGTTCCTAC TTTCTCTGAG
86151  TGCTACTACA GTGGATATAG TGGGAGAATC AACAGTCTCC CCTTAAGGGG
86201  AGACTTTAAG AAGCCAATTC ATGGACCCCT TCCAAACTTG CAGAATCACA
86251  TTACTAAGAC AGAGGCCTGG AATCAGAAAT GATTTATATG TACATTGAAA
86301  CTTGAGAGGC AATTCTTTGC TAAGTGGCTC TCAGCCTAGG TGTCCAATCA
86351  TAATCACATG ACCACTTTAA AGAAATACCA TCACCTGTGC TCTCCCCACA
86401  GGTTCTGTCT GTTGAGCTTG GTGGGCTCAT ACATATAGTT TTAATTAGGA
86451  AGTCAAATTG TCCCTCTTTG TTGATTACAT AATATTATAT CTAGAAAAAT
86501  CTAAAGACCA CCAAAAACTT TTAGATTTGA TAAATAAATT TAATAATGTT
86551  TCAGGATACA AAAATCAATG TAGAAAAAGT AGTAGCATTT TCATACACTA
86601  ATAATGATCA AGCTGAGAAC CAAATTAAAA GGTCAGTTTT TTTTACAATA
86651  GCTACAAAAA AGTAAAACAC ATAGAAATAT AATTACTTAA GTAGGTGAAA
86701  GATCTCTACA AGGAAAACTA CAAAACTCTG ATGAAATAAA TTGTATATGA
86751  CACAAACAAA TGAGAAAAAC ATCACATGTT CATTGATTGG AAGAAGTATT
86801  ATCATTAAAA TGACCATACT GCCCCAACAG TCTACATATT AAGTGTAATT
86851  CCTACAAAAA TTCTAATGTT AGTTTTTATA GAATTAGAAA AAAATTATAT
86901  TTATATGGAA CCATAGAAAA GCCTCAATAG CCAAAGCAAA TTTGATCAAA
86951  GACAACGAAG TTGGACATAT TACATTACGT GACTTAAAAT TATTCTAGAA
87001  GGCTTTAATA ACCAAAACAG CATGGTAATG ATGTAAACAG ATGCACAGAT
87051  CAATGGAGCA GTATAGAGAA CCTAGAAATA AAGCCATATA CCTACACACA
87101  ACTGATCTTT TCCAAAGTCA ACAAAAACAC ACACAGAAAA ATGACATTTT
87151  ATTCAACATA TTGTGCTGGA AAAATTACGT TACTATATGC AGAAGTATGA
87201  AATGGAACCC CTAACTCTCA CCATATACAA AAATCAACTC AATATGGATT
87251  AAAAGGCTAA AATGTAAGAC CTGAAAAGAT AAAAATTCTA GAAGAAACCC
87301  TATGATAAAC TATTCTGGAC ATTGGCCTAG ACAAATAATT CATGACTAAG
87351  ATCTCAAAAG TAGATGCAAC AATAACAAAA ATAGACAAAT GGAACTTAAT
87401  TAAACTGAAA AAGCTCCTGA AAAGAAGCTT TTATTTAATA GGTGAGCAGA
87451  CAAGCTATGG AATACAAAAA AAATGTTTGC CAACTATGCA TGTGACAAAG
87501  AACTAATGTC TAGAATATAT AAAGAAATCA AACATCTCAA CAAGAATAAA
87551  ACAAGAAACT TCATTAAAAA GCAGGCACAT AACGGGAAAA GATATTTTTC
87601  AAAAGAAGAC AATGATGGCC AATAAGCATG TAAAAAATGC TCAACATTGC
87651  CAATGATCAG AGAAATGCCA ATTAAAAACG ACAGTGAAAT ACCATTTTAC
87701  ACCATTCATA ATGGCTAATT ATTAAAAAGC AGAAAAATGA TAGATATTGG
87751  TAAGGATACC GAGAAAAGAG AATACTTATA CATTGTTTGT GGGAATGTAA
87801  CTTTCTACAG CCTCTATGGG AAACAGTATG GAGATTTCTC AAAAAACTAA
87851  AAAATAGAAC TTCCATTTGA TCCAGCTATC CCACTACTGC GTATCTACCC
87901  AAAGGAAAAT AATTCACTAC ATAAAGAAGA TACCCACACT CATATGTTTA
87951  TTGCAGGATT ATTCACAATA GCAAAGATAT GGAGTCAATT TAAATTTATC
88001  TATCAATGAT TGAATAAACA AAATTTGCTA TACATTTATA CCATGGAAAA
88051  CTACTCAGAC ATAAAGAATA AAATTATGTC TTTTGCAGCA CCATGAATGG
88101  AACTGGAGGC CATTATTGTA GGTGAAATAA CTCAGAAACA GAAAATCAAA
88151  TACTGCATTT TCTTACCTAT ATATGGAAGC TCAATAATGC ATACACTTGG
```

```
88201  ATATAGAGAC TGGAAAAATA GACACTGGAG ACTCAGAAAG ATAGGAGGTT
88251  GGTAGAGGGG TTAGAAATGA GAAAAACACC TAACTGGGAC ATGAGCACCG
88301  TTCAGGTGAT TGTTACACCG AAAGCACATA CTTCATCACT CTGCAATATG
88351  TCCCTGTCAT GAAACTTCAT TTGTACTAAC ATATTAATAA AGAGAAAAAA
88401  ACTGACTTTT ATCAAGAGAG CAGAATGAAT AGACCTTCTA CTTTTCATAA
88451  ATACTTAGGC AGAAAAATAA TTTTAATAAA AATAAATAAA AAATGTATAT
88501  TATATATATT TTACATAATT TATATATTGT AAATATATAT CAATATTTTT
88551  ATATATCAAT ATTTGTATAT ATAAAATATA TATCAATATT TTATATATAT
88601  AAATATATAT TATATATTAT ATATATAATA TATTATATAA AATATATATA
88651  ATATAAAATA TATTATATAT TATATATATT TATATATATA AAATATATAT
88701  ATATGGGGTC AGGGTCTCAC TCTGTCACCC AGGCTGGAGT GCAGTGGCAT
88751  GATTTCAGCT CACTGGAAAC TCTGCCTCCC GGGTTAAAGT GATTCTCCGC
88801  CTGCCTCAGC CTCCCGAGTA GCTGGGATTA CAGGCGCCCC CCACCATACC
88851  CAGCTAATTT TTGTGTTTAT AGTAAAGTAC AGTGTTTCAC CATGTTGGCC
88901  AGGCTGGTCT TAAACTCCTG ACCTCAGGTG ATCCATCTGA CTCAGCCTCC
88951  CAAAGTCCTG GGATTACAGG CATGAGCCAT CACACCTGGC CAATAATATT
89001  GCAATATAAG AATGGTATAA AAAGACACTT TGATGAGTTA GAGTAGTTCT
89051  TAGCGCAGAC AATATGCAAG ATTCTAAGCC ATTAGACATT TGTAGACAGA
89101  ATATCTAACA GTGTAAAATA AATAACACGA AGATTCATTG GCAATGAGAA
89151  ATTGACTTTT TTCAATCATA TTAGATGACA TTAAAACCAT TATAAAATTT
89201  ACTGTTTTGT ACATAATAAA GGATCATAAT GTTAAACAAC TTCATTAAAA
89251  GTTTGACAAA TTAGGCATAT AAATAGGCAG CATGTTGACC AGTAAACAGA
89301  AAATACACTT TTCAAAGACC AAACAAAATT ATTTTTATTT ATTTATTCAT
89351  TTATTTATTT ATTATTGGTG GATGAGCAGC TTTATTAGCT GGGGATATAG
89401  TGGGGTCCTC TCCCTGGGAG GTGGGGTCTT TCACTGGTCA CTCCCGGCAG
89451  TGGTCCAGGA GGCGCCAGGC AGTTCAGTGC TGGGCTTAGC TGGGGGCTGA
89501  GCCTTGAAGA AGGCGAACCG TGCAGGGAAG TAGTAGCTGT GGGGTCTCAC
89551  CTCCCGCTCC GCCCTGCTGC ACTGGGTCTC CTGGTGCTCC TCAGGGTCCC
89601  GCCGAGCCTG AGTCTCTATA GGACAGTGGC CCATCCGGCC CGAAACCTTC
89651  TCCTCAGAGC CCAGTTTGAC GCAGGCCAGG CATTTCCACT TCCTTCCCTT
89701  GGGATGGACT TCGCACTCGT GTTTCTTCCA GTCCTTCCGC CGGCCGCTTG
89751  TCTGCCTGAG CTTAAATTCC AGTTTCACAA ATGTTCCAGC TGGGAAGGAC
89801  GTATCCACCG CGCCGTCCAC ACCGTTCTCC CGGAAGGCCC ATGCCGGGGA
89851  GGGTGCATTC CTCCAGGGCT ACCTGCAGGC CCCGGAGCTG GGCCCCGGAG
89901  CTCGGACCCG CCTGCCTCCC CAGCGCCCCC CGCGCCCACC CACGGGGCCA
89951  GCAGCATCCG CAGCCGTGGC TTGCTTCTGC GGTCTCTCAC TCTGGCCCTG
90001  CGAAGCTCCT GTGCACCGCT CAGCTCTCTG AGCCCGCTGG GAGGTGCCTC
90051  CTCCCCTGCT CTTCCCCTGG GTGGCTATGC CCACAGAACT CTGGGCAGAG
90101  GTCAAAGAGC CAGGAAATGT CTCTTTCTCC AAATTGACTT TGGTGTGCGC
90151  CTGGTTCTCT CCACTCCCTC CTGCCCTGTC CACACTGTTC CCTGGGGCCC
90201  GCAGGTTTAG CAAAGTTCCC TGCCCCCTGC CTGGGCCAGG AAGCAGTCCT
90251  GTTGCCCACT CCCACCCTTC AACCCTTTTA TGGCCCATTC TCTCTCCCCA
90301  CTGGGTCTCC CACACGACAA CCCCTCCTCC CTACTGTCCC CGGAGCCCCT
90351  CTCTGGTTCT CGCGCTCAGC CTCTTCCCTG ACTGCTTCTC CATCTCCATC
90401  CTGAATCTCC CAGCTCCAGT AGGGTGCCCC CCAATCCCAG GGCCCAGGCA
90451  AAACCAACAA AATTATTTAA AATGGGAATA TTTGAATTCC ACTGAATTCG
90501  TAAAAGCAGA AACCCATCTG GTTATATTTT AAAGAATTAT GAATTAATAA
90551  TAGCAACTAT CAATTTAAAG TGTAACCAGG GTTAAGAATA CCCACCATTT
90601  TAACAACAAC AACAATGAAA GCGTGTTATA GCTAACAAAC TAGTTCTGGA
90651  AAGTTGCAGG ATACAATATT AACATGAAAA TCAGTTGCAT TTGTATACAG
90701  TAACAACAAA ATATCTGAAA AAGGAATAAA GAAAACAATT CCATTTACAA
90751  TATTATCAAA TAGAATGAAA TACTTAATTC ATTAAATAGA ATGAGTTTAA
90801  CCAAGAAAAT TAAAGATCTG CATACTGGGT CGGGCGCGGT GGCTCATCCG
90851  TGTAATCCCA GCACTTTGGG ACACCAAGGC GGGCGGATCA CGAGGTCAGG
90901  AGTTCGAGAC CAGGCTGTCC ACATGGGGAA ACCCCGTCTC TACTAAAAAT
90951  AAAAAAAAAA TTAGTCGGTG GTGGTGATCT CCTGTAATCT CAGCTACTCA
91001  GGAGGCTGAG GCAGGAGAAT CACTTGAACC CAGGAGGCAC AGGTTTCAGT
91051  GAGCCAAGAT TGGGCCACTG CACTCCAGCC TGGATGACAT AGTGAGATTC
91101  CATCTCAAAA AAAAAAAAAA AAAAAAAAAA ATCTGCATAC TGAACACTAT
91151  GAAATGTTGA TGAAAGAAGT AGAAGAATAG GCCAGTTGCG GTGGCTCACG
91201  CCTGTAATCC CAGCACTTTG GGAGGCCGAG GTGGGTGGAT CACGAGGTCA
91251  GGAGATTGAG ACCATCCTGG CTAAAACGAT GAAACCCCGT CTCTACTAAA
91301  AATACAAAAA ATTAGCCAGG CGTGGTGGCA CGCGCCTGTA GTCCCAGCTA
```

91351 CTCGGGAGGC TGAGGCAGGA GAACTGCGTG AACCTGGGAG GTGGAGCTTG
91401 CAGTGAGCTG AGATGGCGCC ATTGCACTCC AGCCTGGGTG ACCGAGCGAG
91451 ACTCCGTCTA AAAAAAAAAA AAAAAAAAAA AAAAAAAAGC AGCAAGCAAG
91501 CAAGCAAGCA AGCAAGAAAG AAGTACAAGA ATACGAAATG TGAAATATAT
91551 CCTGTGTTCA TGGATTCTAA AAATTAATAT TGTTAAAATA TCTATACTGG
91601 ACAAAGTCAT CAACAAAGTT AAAGAAATTT CTATCAAAAT TTTAATGCCT
91651 TTAAAATAAG TGTAGAACAA ACAATTCTAA AATTAGTATA GAGCCATGAA
91701 AGACCCCAAA TAGGCAAATA CTGTGAAGAA CAGAAAGGCT GAATGCCTCA
91751 AACTTCCTGA TTTCAAACTG TATTACAAAG CTATAGTCAT TAAAGTAGTA
91801 TAGAACTTAC ATAGGAACCA CTGAAACAGA ATAGAGGACC TAGAAATAAA
91851 TTCACCCATA TGCAGTCAAC TGGTCCTACA GAACCAGGAA AAGATAGGGA
91901 CATCAAGAAG TGGTTTAGAA AAAACTAGAT ATGCACACAC AAAAAGTGAA
91951 ACTTTCTTAT ATCATCACAA AAAATGAGTT TAAAATTAAA GGCTTAAACA
92001 TAATAACTGA AATCACGAGT CGTCTTTAAA AAATAGGGAA AAAGCTCCTT
92051 CACCCCGGTC GTGGCAATGA TGTTTTGGAT TCTACACAAA GAACACAGGC
92101 AACAAAAGCA AAAATTTAAA AAATGGAACT ATATCAAAGT TTCTGCATGA
92151 TAAAAGAAAG AATCAAGAAA ATATAAAGAC AATATATGGG ATGGGAGAAA
92201 ATTTTCGTAA ACCATATGTA GGATAATATG TTGCTATTCA AAATATACAG
92251 AATACTAATC AATATGAAAA AAGCATCCAC AGGAGAACAA AACAAAAACC
92301 AATTCCCTGA TTAATTGGGC AAAATATTCA TTTTTCCAAA GACATACAAA
92351 TGGCCAGCAG GTATATGAAA AGTTTCTCAA CATCACTAAT TATCAGTGTA
92401 ATTAAAATCA AAATCAAAAT GAGAAATCAC CGTATCATGG TGTTAGGATA
92451 ACTATTATCA AAGTGTCAAA AGAACAAAGT GTTAGGGTGC ACAGAAAAGA
92501 CAATATTTGC ACACAGTTGC AGAGCATGTC CTTTGGTGCA GCCATTACAA
92551 AAAAAAAATC CAGTATGGAG TTTCCTTAAA ATTTTTAAAC TTGAAGAACC
92601 ATTAATCCCA ATTTGGAGAA TATAGCCAAT GGACATAAAA TTTAAGTATA
92651 GCCGAGGGCG TTTGCATGCC TCTGATACAA ATAGATGGAT AAACTGTAAG
92701 ACAGAGATAA TTTCAGCTTT AATAAGGAAT GAAATTGTTT TAGTTACAAA
92751 AATATTGATG AACCTTGAAG ACATGATGCT AAGTGAGATC AGCAAAATAC
92801 AGAAAGGCAA ATATTGCATG ATCTCATTTG TATGTAGATA TATTAAAAAA
92851 AAGAAAGAAA GCGGGAGAAG GGTAGTTTGC ATGGGCTAGG AAATGGGGAA
92901 AGGAGGAGAT ATATCCATTG AAGGGTGCAT ACCTTCAGTT ATATAATGAA
92951 AAACTGCTGG GGACCTAATG TGCAGAATGG TGACTATAGT TAATAATAAC
93001 GTGTAGTTGA AATCTGTTAC TAAAGTAGAC CTGAGGTGGT TTCACTACAC
93051 ACACTGAAAT GTATAAAGTA ACTATGTGAG GTGTTAGATA GGTTAACCAG
93101 CTTCACTGAG ATAATTTAAA GACGTATACA CATCTCAAAG CACTCTATTG
93151 TATACCCTAA ATAAATACAC TTTTCAATGT GTAAATGTGG AAAAAAATTA
93201 ACAGTAGTCA GCAGGTCACA TCTAGCCACA TGTAAACTTT ATTTAAAATT
93251 GTTGGCCACC CTTTCTGGCT CAACCCGGGA ACAACCGGAG CACTTCTGGC
93301 CCCTTGGACT TTGACGCTCC TCCCACTGTT CCTGTACTGG GGAAGATATG
93351 GATAGTTCCC AGGGGTGGGT GGGCGCAGTG GCTCACGCCT GTAATCCCAG
93401 CATTTTGGGA GGCCAAGGTG GGCGGATCAC TTGAGGTCAG AGGTTGAAGT
93451 CCAGCTAGAC CAACATAGTG AAACCCTGTC TCTACTAAAA ATAGAAAAAT
93501 TAGCTGGGTC TGGTAGCCTG CGCCTGTAAT CCCAGCTACT CAAGAGGNNN
93551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
93601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNGTG
93651 TGAGGCAGGA GTATTGTTTG CACCTGGGAG GCAGAGGTCG CAGTAAGCAG
93701 AGATCGTGTC TGTGCTCTCC AGTGTGGGCC ACAGAGTGAG TCTGCGCTTT
93751 AAAAAAAAAA TATAATATAT ATATTATATA TTATATATAA TATATATATA
93801 ATATATTATA TATCTAATAT ATATATTATA TATAATATAT ATTATATATA
93851 TACTATATAT ATTATATATA TAATATATAT ATAATATATA TAATATATAT
93901 AGTATATAAT ATATATATTA TATATATAAT ATATATTATA TATAATATAT
93951 AATATATAAC ATATATATTA TATGTTATAT ATATATGTAT GTATGTATTT
94001 CCCAGGGTTT TTTCACATCC CAGACTGTGC AGATCAGGGA GGGCCATCAG
94051 CAGTCCCTTT ACTAGTTGCC CACAGACCCG TTTGGCTCCA GGCAATGCAT
94101 ATCAGGGCAT CCAGGTGGGA GCCACAGCAG CCAGCAGGGA GGAGGCTGCC
94151 TTTGGCTGCC TGCAGTCTGG TGGCCTCTGG CTCCGCAGGT GTCTGTAAGG
94201 CCCCAGCGCA GCCCCTCCTC ACATTGCCTG TGAAGGCAAT GTAACTGCAG
94251 AACATACTGA GGAGGAAATT GAGGAAAAAA ATGTCCCCCA TATATTAGAA
94301 TAATTGTCAT TAGATCGCTT ATGCTGCAGA CAGGGTCACT GTTTATCATT
94351 ACTGCTGTGA AACCCGCTAC AATTGGAAGA GAAAATGATC TAGGGCATAC
94401 CCTTTGATTC CCATAAGTTG GTGTCAGTAG GTTTCATGCA GAAATTTATT
94451 TTGACCATGA TTTACAAAGT TACGTCCTTC TAGATCAAGG CAGTCAAAAT

94501 TGAACAACTG TTCATGGAAA ATGAATTGTT CAGCTGAAAA CTAAATGAGA
94551 CCCTTATGTA CGTGAGCATA GACATAAATT TAAAATTGGA GAGACTGTGT
94601 CACCCTTTCA CATTCACCCT GGCAGTGATA GCTGTGATGG CTGTGAACCA
94651 TGGCAGGTTA GAACTCACTT TTGCTTTGAT AAGAAAGATG AATCTTTTGT
94701 TGGTCTACCA CTAAGTAAAG AGGAAAATGA GTTGGAAATA AGTAAACAAT
94751 TAAAGAAAAT ATGACTAAAA TGGGGTTTAC AGAATATAGA TTATGAACAT
94801 GAATAGACAC TGAATAATCT AAAATATAAA GATAGAGCTG GAAATCATAG
94851 TGAGCAGATT GGAAGTGAAG GAACATTCTA GAGAGATGTT GCTCCTGCAT
94901 CTGTTCATTC TGAAATTACT GACAGTAACA AATGTCAGAA GATGTTGGAG
94951 AAGATGGGTT GGAAAAAAGG AGACGGCCTG GGGAAGGATG GTGGGTAATG
95001 AAATCTCCGA TTTAGCTTCA GCTTCAGTGA ACATATGGAG GCTTGGGGAC
95051 AGGCAATCCA TCCTCAATTG AAGACGTTCA GCTTCTCTGA AACAAAAAAC
95101 AACCTGGGAC AAAGCGTGAG AGAGGTTTGC TGAAAATTTC CCAGAAACCA
95151 AACCTTAAAA AGATGACCTA GGTACCAGGC CTTGGATAAA AGGAACTGTA
95201 CAGTGAAGGT TAATCATAGA AGAAAACTCA AGCTTTTCAA AAAATAGAGT
95251 TTGGAAACTC TTATTTTATT ATATATTTGC AGTACTTTTC TCCCCAAAAG
95301 AGTCTGTGGC ACAGAGGAAC AGTGTCACAG TTTACCCCTT CTTGATTCAG
95351 AAATGTGTAA TAAAGTTTGG TTTGCAAATT TTAAAAAACA TTTTTTAAAC
95401 TAATAAATAT TGACTCAAGT TATTCAGTAA GTGGACTAAA GTTTACAGGT
95451 TAAAGATGAG TTTATCAAAC TTCCTTATTT TATCTTGTCA TTTATGACAT
95501 CCATATAAGC AAAAAGCCAC ATAAGCAAAA CTCATAACCA CTAATGACTT
95551 AAATGTACAT TTGTCTGTGT GTCCATGTAT TCACAGTAAG GTGCACAGCA
95601 AAAGAAACAA CAAAAGTTTA TAAAAATAAA TCTGACTACA TGCATCATTG
95651 TTTATGCCCT TTAGAAACTA GATAAAAGAA CCTCTTATAC TTGAAATAGC
95701 CTAAATATTA TTGAAGAACA AATGAACAGC TGATATATTG TAGAAAATTA
95751 GTCAGTGTTT TCTTTTTTGA AGAATCCGTT TATTAGTACT ATATCTTCAG
95801 TATTTATATT GGTTTGTTTC ATAGCTAATG TGATATTTAG ATATGAACAA
95851 CTGAGTACAG TGTTGAAATA GTGTGCTGGC ATTTGTAGTT TTCATAAATA
95901 TTATTGCAGG CAGTGGTGTT GTGCCACAGA AATCTGATTT CTAGTACATA
95951 AGGATTACTT AGCCAGGGCC TCATGTTTAA GATATTTAAT GAAAATGTCT
96001 TCAACTGCAA TAAAAACATT GTAACATTAA AAATACTCTT TTCAATAAAT
96051 TCTAAATTAA AAAATTCAAA TGATACCTTT TATGGAGTTA GGGAAGTGCT
96101 AATAAGGTAA AATGGCAACT GAAGCCAAAA AATGTAAATT CAGGTAAATA
96151 CTTTTACCCT TATTAGTTGT ATGTTAAAGG AAACAATAAT AAACAATAAT
96201 TCTGACCATA TACTATTTAC TGCAGTAAAG TATTTGAGAA ATTTGTGCAT
96251 AATACATAAT TAATTTTCTA ATGGTATAAA AGTAATCACA TTCTACAAAT
96301 TATTACAACA CGGTCTATTG AAGACAGCGG CATTTCAAGT GAAGAATCTT
96351 AGAGTTTCTC ACGAGGCAGT GATATATGCC ATATGAAATC TAGGTAAAAT
96401 ATTTTACTCA TGTATGCAAC AGTTGATATT TCTGCTTTCC AGGAAAATAA
96451 CATGCTTTAA AAACTTAATG CAGGTAACTA AAATCACCAA GAAGTTATAA
96501 GAACTCACAC AATGACTAAT ATAGTTGAAA GAAAATTATG AAAACTTCTC
96551 AGGACCAGAA GTAAATAAAG GGTATAATCC AGAGAAGTAT AGACCTAAGA
96601 GGCCAGTGCT CCATTCAGAT GCATCTGATC ACAAGGACAT CAGATCCACA
96651 TGGGTTGTTC AGCTTCTGAC AAAGCAAATG GAATAAACAA AGAGAAACTG
96701 CCTGCAGGCA TCCAGAGTGT GATGTCTCCT ATATGTGAGA GCTAAACTAT
96751 AAATTATGTT GCACACAAGG AGAAGAGCTA TTAGAGTGAA GCATTAAATT
96801 ATATGGCACA TAAAACTCAG ATATGAGGAT TATAATTACA TGTTTGCCAA
96851 TATAATGAAA ACACAAGAAA CCGAAATAAT GTGGAAAAAA CTATGAGCTT
96901 GTCAAGTCAT ATTTGGAAAA GAGGCGAATG GAAAGTAAAC TTGAAAATTT
96951 GGTAAAACAA CTTAATGTAC AGCATACAAA TTACACATTG AAATAGACTG
97001 AGATGGTGAG AGGACTAGTA AACTGGAATG CTGAGCACAA TTAATGTAGT
97051 AATATACTTT TTAGAAGGCA AATTAATACA AAATACAAAT ATATATGAAA
97101 ATTAGATGAG AAGAAATGAA AAACATTTAA TTGCTTTACT ATACTATATA
97151 AACAGGAGGG CAATATTCAA AAAAATCAGT GACTCATAAT TTTCAGAAAT
97201 TGGAAACCAG GCATGAATCC TATAAAAGTT TAGAGTGTGA TGTGTCAGAA
97251 AAGATAAATT AAGAAATACT TAAAATAAAT AATGATTTTG GTTATTGTGA
97301 ATACTGCTTC AATAAACATG GGAGTGCAAT TATCTTTTGG ACATACTGAT
97351 TTTATTTCCT TTTGATATAT ACCCAGTGGT GAGATTGCTG GGTCATATAT
97401 ATGGTAGTTT TATTTATAAC TTCTTGAAAA ACCTCCATGC TGTTTTTCCT
97451 AATGGTGGTA CCAATTTAAA TTCTCACCAA CAGTGTATAA GTGCCTATTT
97501 ACAGATGAAT GAAGAAAATG TTATACATAC ACACAGGCAC ACATACACAC
97551 ACAGGAATAA TGCTCAGTCA TAGAACAGAA TGAACTTCTG TCATTTATGT
97601 GAACATGGAT GAACCTAGAG GACATTATGT TAAGTGAAAT AAGCCCATCA

```
97651  CAGAGAGAAA AAGACCACAT GATGTCACTC ATATGCAAAA TCTAAAAAAG
97701  CAGATCTAAT AGAAGTAAAG AGTAGAATGG TGGTTACTAG AGGCTAGGGA
97751  GAGTGATGGG ATGAAGGAAT GGGGAGAGGT TTGTCAAAGT TACAAAGTTA
97801  CAGTCAGGGA GAATAAATTC TGATGTTCTA TTACACAGTA GGGTGACTAT
97851  AGCTAATATA ATGTACTATT TATTATAACA CAGCTAGAAG AGAGGTTTTT
97901  GAATTTTAAT AGCACAAATA AATTATAAAT GTTTACACTG CTAGATATGC
97951  AAATTACCCT CATTGGGTCA TTTTACAGTG TAAACATGCA CTGAAACATC
98001  ACACTGTAGC CCATAAATAT GTACAATTAT TATGTGTCAA TTATACATAA
98051  AATAAATCTT AAAAAATAAT AATGATCGAA TTGCATAAAA TCAAAACGAA
98101  GAAAAACTCT TAAAAGTAGT GAGATAGAAG ACAGCTAACC TGTGAAGCTG
98151  AGGGAAAACA CTTACCTCTG TATCAAGTCA GAAAACAGAA AAATTAATAT
98201  CAGTAAAATA TTAAAAGCTT CTGGACAGCA CAGAAATCTA TTAATAGAGT
98251  GAAAAGACAA CATATAAAAT GAAAAAAAAT ATGTGCATCT GGCAGGGAGT
98301  TAATATCCAA AATATATATA TGACTCAACT CAACAGCAAT AAAAATCAAA
98351  AACCCAACTA AAGAGTGAGC AAAAAACCTG AATAGACATT TTTCCAAAGA
98401  AGACACAAAG AAATGCTCAA CATCACGAAT CATCTGGAAA ATGCAATCCA
98451  AAACCACATG AAATAGTACC TCACCCCATT TAGAATGGCC ATTATCAAGA
98501  AGACAAACGA TAACAAGGTT TGGATGAGGA TGTGGAGAAA AGGGAACACT
98551  TATACACTGT TGGTGTGAAT GTAAATTTAG TACCTTTTTA GATTCATTTA
98601  TTTATTGATA AGAAACCTGC ATGGAAAACT ATGCAGGTTT CTCCAAACAA
98651  AAAACAAAAA CCTAAAAATA GAACTACCAT ATAATCCAGC AATTCCATTA
98701  TTGAGTATAT ATCAAAAGGA AATAAAATTA GTATATTGAA GAGATAGCTG
98751  CTCTCTCATG TTTATCATAG CTCTATTGAC AACAGCCAAG ATATGGAATC
98801  AATCTTGTGT CCATCAGGGA TAAATGGATA AAGCGAATAT GGTGTATACA
98851  CATAATGGAA TACGATTTAA CCATAGGAAA AAAAAAATCC GCCCTTTGGG
98901  AGGCCGAGGC GGGCAAATCA GGAGGTCAAG AGATCGAGAC CATCTTGGTC
98951  AAAATGGTGA AACCCCTTTT CTACTAAAAA TACAATTAGT TGGGCGTGGT
99001  GGCACGCGCC TGTAGTCCCA GCTACTCAGG AGACTGAGGC AGGAGAATCG
99051  CTTGAACCCG GGAGACGGAG GTTGCAGTGA GCCAAGATCG CGTCACTGCA
99101  CTCCAGCCTG GCGACAGACG TTCCGTTTCA AAAGAAAAAA ATAATATTAA
99151  TAAAAAGAAT AAAATCCGGC GCTGCGCGGT GACATCAGTC TCTGTCGTTA
99201  ATGCCTCGCG CCGGCTACCG TCCTGCGCAG TCTCTTTCTG AGGACCCCCC
99251  CCCCCACTCT CCGCCTTCCA ATAAGGAGTT CAGGTTTTGC GGTCGCCGTG
99301  GTTGCTGTTC CTGCTGCCAC AGGTTGGAAC TGGAGATGCC TCTTCCTTCT
99351  CTCAGGACAG AACCATGAGC CTAGCGGCAG CGCCGGTTCG CGAAGCTCCC
99401  CCTCCGCCAA CGGGCGCCTC CTCAGAGCCG TCCGTGCCCG CCCTGCCGGG
99451  AGCTGACCCG CAGCGCAGTG CAGAGTTGCT CCTGTTGGCG GTGACCAGGG
99501  AGGGACTGGA GCGGCGGATC ATCTCCAGGA AGCGGGCTGA GTAGGAACTG
99551  CAGCCGCCAC ATCCTCTCTT TACCCGGGGA TGTGCAGGAT TACCGTGAAA
99601  TCATGACTCG TCATCCTGCG AATTACCAAT GGGAAAATTG GAGTCTAGAA
99651  ATTATTGCCT CCATTTTAGC CCACCGGTTC CCCAGTAGCT GTATTGGGGT
99701  GATGAAGTGC TCCGGAACAA ATGCGCTGCC ATGATAGTTT TCTGAAAAGT
99751  AACATGTTTG GTTTCCCAGA ACACAATACA GACTCTGGAG CTTTTAAGCA
99801  CCTTTATATG TTATTAGTTA ATGCTTTTAA GTCAGAGTAG TTTATCAAAG
99851  GAAAATTTGA ATGATTGGAA TAAGGACTCC ACAGCATCTA ATTGTAGATG
99901  TCCAATTCTT CTCATACTAC AAATCATTTC CAGGAAGGAA AAGATAGGAC
99951  CTTTGAAAAA TCTGATGGAT CGGCCATGTG TTTTTATCCA CCATCACTAA
100001 ATGATGCATC TTTTCCTTTG ACTGGATTCA ATAAATGTGT TGTTTTGAAT
100051 CAGTTTCTTT GAATTGAAAG AAGCCAAGAA AGACAAAAAC ATAGATGCTT
100101 TCATTAAAAG CATAAGTACA ATGTATTGGC TGGATGGTGT TCATTCCGGA
100151 GGAAGGAATA CTGGAGTTAC TTATCCACAA GTCTTGAAAG AATTTGCACA
100201 AACAAGAATT ATTGTTCACA CCCATGGAAC ACTTTACCAA GTATGTGATC
100251 TAATGAGATC TTGCATTGGA AAGGAGCAAA ATAAATTTTT TAGATACTTG
100301 GGGATATTGG TATGCTGGTG AGTAGCTGAA TTCATTTCAC GAAGGAAGCT
100351 CCCTACATAG AGAATCCCTT TCAGAATTCA TGAAGTGTTT GAGACTACAA
100401 GTATATTAAT GTACTTGTTC AGCGGAAGAG CATAAGCACT TTGAGTGTTA
100451 TAAATTCAGA TAATGGAATG TACTTCATAG ATGTATTGTC AGTTTGGGGG
100501 TATGGAGGGA AGCACACATT CCTGAAAAAT GAGTGTAATG TGC
```

FEATURES:
Start codon: 1000
Exon: 1000-1062
Intron:1063-2440
Exon: 2441-2514
Intron:2515-3016
Exon: 3017-3171
Intron:3172-33878
Exon:33879-33959
Intron:33960-39894
Exon:39895-40002
Intron:40003-50367
Exon:50368-50398
Intron:50399-50736
Exon:50737-50886
Intron:50887-53571
Exon:53572-53605
Intron:53606-65379
Exon:65380-65613
Intron:65614-74191
Exon:74192-74299
Intron:74300-79441
Exon:79442-79459
Intron:79460-99355
Exon:99356-99544
Stop codon: 99542

336000014324295

SNPs:

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
948	T	-	Beyond ORF(5')			
1149	G	A	Intron			
4199	G	C	Intron			
4352	T	G	Intron			
6493	A	G	Intron			
14047	T	G A	Intron			
14136	G	T	Intron			
14238	G	A	Intron			
14260	T	G	Intron			
25736	C	T	Intron			
26321	G	A	Intron			
31359	A	G	Intron			
35098	T	G	Intron			
40532	C	G	Intron			
41706	A	G	Intron			
51095	G	C	Intron			
53101	G	C	Intron			
54556	G	C	Intron			
61872	-	T	Intron			
62172	G	C	Intron			
62860	A	C	Intron			
67086	C	A	Intron			
67621	A	C	Intron			
70582	A	T	Intron			
74175	C	-	Intron			
74478	T	C	Intron			
77092	T	G	Intron			
77328	T	C	Intron			
77385	G	A	Intron			
77947	C	T	Intron			
79395	G	C	Intron			
81111	-	C	Intron			
81610	G	A	Intron			

Context:

DNA Position

948

TGTGTCCTCCTGAATCTTAGTGGCCTTCTAAAGGCGGGTGTGATCAGCCATGGGTATCAG
AGACACTGGAGTCCAGTAGCTGCTAGGTGGGACACGGGCACAATTTCACTTGCAGACCAG
CTGCACGGAGTGGATAAAGAGAGAGTTCTGTGTGGGAATCTCCTTTGGTGGATCATCAGG
GAGGTGAAGTCTTTGTCATAGCCTCATATCCAGCTTGTGTGATACCAATTCCAGTGAAGC
TGGAACAAGCTGGCACTGCTCAAACAGGCCTACCAAGACATCATGTTTTTTTTTTTTTTT
[T,-]
CCACCAAACCTGGACCTGAATGGGGATGTGGACACACATAGAGTCCAGAGGATGGGACCC
TGGTCAGTGGTGGTGCTGGTGTGCGGCATGAAGCAGCTGGGGCAGGCCCTCCAGGTGGGA
GGAGGAGCCAGCCTCTCCTGTGGGGTCCTGGGCAAGTCATTTCCCTCCTTGACGCTCTGT
TTCCTCATCTGGAAAATCCAGGAAGCCTGTTGTGCAGTCCTCAGAGGGTCATGATAAGGT
GCAAAGGAGGAAGAAATTTTGAGAGTTCTTGTGCCTTCCTCTTCCTGAGAGGGATGATGG

```
1149        CCTCATATCCAGCTTGTGTGATACCAATTCCAGTGAAGCTGGAACAAGCTGGCACTGCTC
            AAACAGGCCTACCAAGACATCATGTTTTTTTTTTTTTTTTCCACCAAACCTGGACCTGAA
            TGGGGATGTGGACACACATAGAGTCCAGAGGATGGGACCCTGGTCAGTGGTGGTGCTGGT
            GTGCGGCATGAAGCAGCTGGGGCAGGCCCTCCAGGTGGGAGGAGGAGCCAGCCTCTCCTG
            TGGGGTCCTGGGCAAGTCATTTCCCTCCTTGACGCTCTGTTTCCTCATCTGGAAAATCCA
            [G,A]
            GAAGCCTGTTGTGCAGTCCTCAGAGGGTCATGATAAGGTGCAAAGGAGGAAGAAATTTTG
            AGAGTTCTTGTGCCTTCCTCTTCCTGAGAGGGATGATGGTGAGAACAGTGTGGATAACCA
            CATGCAACTGAGTCCCCAAAAGGCCCTGTGAGGAAGGTGTTGTTCCCATCATGCTTCCCA
            GATGAGGAAACAGTCTCAGGGAGGCCTGGCTGCATGCCCAAGGGTTACACACACACTGAA
            TGTTGGAGCTGGGAGTAAATCTAGAGTCAGGGCTCACTGGAGGTGGTGGGAGCATTGCAC

4199        AATGTGGATGATGGTTGCAGAACAATCTGAGCATGATTAATTCCTCTGACTTGCACGTTA
            GAAATTGTTAAAATAGTTAATTTTATGTATATTTTACCACAATGTAAAAAAGGAATTTTT
            AAATGAACAGACTGTAGATACATGCAACAGCATAAATGAATATCACAAATATAATCTTGC
            ATTTAAAAATTGATGTAAAAGTATCCAAACTATATAATTTCATTTATACTAAATCCAAAA
            ATCAAAACTGACATTCTTGCTTTCACTAATGGGAATTAGCTAGTTAAACTAACACTCTCA
            [G,C]
            AGAGAAAAATGATGAATCCTAGATAAAATAGTATATATCATTATAAACACTTCTATATAT
            AATACATATATGAGATATGTGTGTATAAGAAGTGAATGAGGATGTCCCCTGTGCCCTCCT
            TAGGAGAGACAAGAATTGAAGTTATAATCCAGGCCAATTAGCACTCTCTTTAAAAATCAA
            CACTCTTCAAAGGGACACAACAGAATCCAGAGTCTCTATAACTCTTGTATACAGTCTCTT
            GTACACAATTTTCAAATTCGTGAGATGGGTGAAGACACATGAAAATGCAATACATACACA

4352        AAATGAATATCACAAATATAATCTTGCATTTAAAAATTGATGTAAAAGTATCCAAACTAT
            ATAATTTCATTTATACTAAATCCAAAAATCAAAACTGACATTCTTGCTTTCACTAATGGG
            AATTAGCTAGTTAAACTAACACTCTCAGAGAGAAAAATGATGAATCCTAGATAAAATAGT
            ATATATCATTATAAACACTTCTATATATAATACATATATGAGATATGTGTGTATAAGAAG
            TGAATGAGGATGTCCCCTGTGCCCTCCTTAGGAGAGACAAGAATTGAAGTTATAATCCAG
            [T,G]
            CCAATTAGCACTCTCTTTAAAAATCAACACTCTTCAAAGGGACACAACAGAATCCAGAGT
            CTCTATAACTCTTGTATACAGTCTCTTGTACACAATTTTCAAATTCGTGAGATGGGTGAA
            GACACATGAAAATGCAATACATACACAAGATAAAAGGCAGGCAGTAGACATCTCCAAGAT
            ATCCAAGATGTAATCAGCAGACAAGAATTTGAAGGCAGCTATTACAAGTATGCTAATGGA
            GGCAAAGGAAAAAATATACTTATAAAGGAACAGATGTGGAACCTCAGCAGAGAAATAAAA

6493        TCTTGAAACCAGGAGTTCGAGACCAGCCTTGGCAACATAGTGTGAGCACCCTCCGCCCCC
            CAACCTTTTCTACAAAAAAAAGAAAGAAAGAAAAAATAGCCAGGCATGGTGGAGCTTGTC
            TGTGGTCCCAGCTACTTGGGGGACTTAGGTGGGAGGGTCATTTAAGCCTGGGAGGTAGAG
            GCTGCAGTGAGCTGAGATCAGGCCACTGCATGCACTCCAGTCTGAGTGACAGAGCGAGAT
            CCTCTCTCTCTATCTCTTCCTCACTCTGTGTGTGTGTGTTGGGGAGAGGGGTGTGTGTTT
            [A,G]
            TGGGTGTGTGTGTGAGTGTGTATGTGTGTTTATTATTCAAAATGAAAACAACATAATGAC
            AATTATTTTTATTTTTATTTTTTTGAGACAGAGTCTCACTTTGTCAACCAGGCTCCAGTG
            CAGTGGTGCGATCTCGGATCATTGCAACCTCCGACCCCGAGATTCAAGCAATTCATCTGC
            CTCAGCCTCCTGAGCAGCTGGGATTACAGGTGCCCACTACCTTGCCTGGCTAATTTTTGT
            ATTTTTAGAAGAGGCAGGGTTTTGCCATGTTCTCCAGGCTGGTTTTGAACTCCTGAGCTC

14047       TGGAAGCCTCTGTCCAAATCTCCCAGGCTATCTATGTGTGGGAAGTGGAGTCAGGAAAGA
            CCCTAGAGCTTTGGAAGTGATCATTAAGAGAAAACAAAATCCCTGTAAATTAGAGTACAA
            GGGAACATATTCATTAGCCACCTTTAGAGTCAGATGCTCAGGGCTGAGCTGGTGTGGGGG
            AGTGTTCAGACCTGTGCATGCCTGGGAAAGCCTCCGATCATGTATGAATGGTGAGGCTTC
            TGGGTGCATGAGATTGAGTGCCCTTCCTTGGCAGAGCACCACTGAGTGAACAATTGATTT
            [T,G,A]
            GGATCAAGCATATGGAGATCAACTTTATTTTGAATATCTCAAAGACAGAAACTGAAAAGA
            TTTTTTTGCACTTTGAAATTGAGTAAGGGTGTCAGAAACTTCAGTAAAAAAGTCACAGGA
            GGAAACCCAGAAGTCTTATTCATCCCCAGAAACCACCAAAATCCTGATCCAAACTGTAAG
            AATTCACCTCGTAAAAATTTGATTTTTGAATAGGAACAGCTCCGGTCTACAGCTCCCAGT
            GTGAGCGATGCAGAAGACGGTTGATTTCTGCATTTCCATCTGAGGTACCAGGTTCATCTC
```

14136 GAAAACAAAATCCCTGTAAATTAGAGTACAAGGGAACATATTCATTAGCCACCTTTAGAG
TCAGATGCTCAGGGCTGAGCTGGTGTGGGGGAGTGTTCAGACCTGTGCATGCCTGGGAAA
GCCTCCGATCATGTATGAATGGTGAGGCTTCTGGGTGCATGAGATTGAGTGCCCTTCCTT
GGCAGAGCACCACTGAGTGAACAATTGATTTAGGATCAAGCATATGGAGATCAACTTTAT
TTTGAATATCTCAAAGACAGAAACTGAAAAGATTTTTTTGCACTTTGAAATTGAGTAAGG
[G,T]
TGTCAGAAACTTCAGTAAAAAAGTCACAGGAGGAAACCCAGAAGTCTTATTCATCCCCAG
AAACCACCAAAATCCTGATCCAAACTGTAAGAATTCACCTCGTAAAAATTTGATTTTTGA
ATAGGAACAGCTCCGGTCTACAGCTCCCAGTGTGAGCGATGCAGAAGACGGTTGATTTCT
GCATTTCCATCTGAGGTACCAGGTTCATCTCACTATGGAGTGCCAGACAGTGGGCGCACG
TCAGTGGGTGTGCGCACCGTGCGCGAGTGAAGCACGGTGAGGCATTGCCTCACTCACGAA

14238 CTGTGCATGCCTGGGAAAGCCTCCGATCATGTATGAATGGTGAGGCTTCTGGGTGCATGA
GATTGAGTGCCCTTCCTTGGCAGAGCACCACTGAGTGAACAATTGATTTAGGATCAAGCA
TATGGAGATCAACTTTATTTTGAATATCTCAAAGACAGAAACTGAAAAGATTTTTTTGCA
CTTTGAAATTGAGTAAGGGTGTCAGAAACTTCAGTAAAAAAGTCACAGGAGGAAACCCAG
AAGTCTTATTCATCCCCAGAAACCACCAAAATCCTGATCCAAACTGTAAGAATTCACCTC
[G,A]
TAAAAATTTGATTTTTGAATAGGAACAGCTCCGGTCTACAGCTCCCAGTGTGAGCGATGC
AGAAGACGGTTGATTTCTGCATTTCCATCTGAGGTACCAGGTTCATCTCACTATGGAGTG
CCAGACAGTGGGCGCACGTCAGTGGGTGTGCGCACCGTGCGCGAGTGAAGCACGGTGAGG
CATTGCCTCACTCACGAAGTGCAAGAGGTCAGGGAGTTCCCTTTCCTATTCAAAGAAAGG
CATGACAGATGGCACCTGGAAAATCGGATCACTCCCACCCGAATACTGCGCTTTTCCGAC

14260 CCGATCATGTATGAATGGTGAGGCTTCTGGGTGCATGAGATTGAGTGCCCTTCCTTGGCA
GAGCACCACTGAGTGAACAATTGATTTAGGATCAAGCATATGGAGATCAACTTTATTTTG
AATATCTCAAAGACAGAAACTGAAAAGATTTTTTTGCACTTTGAAATTGAGTAAGGGTGT
CAGAAACTTCAGTAAAAAAGTCACAGGAGGAAACCCAGAAGTCTTATTCATCCCCAGAAA
CCACCAAAATCCTGATCCAAACTGTAAGAATTCACCTCGTAAAAATTTGATTTTTGAATA
[T,G]
GAACAGCTCCGGTCTACAGCTCCCAGTGTGAGCGATGCAGAAGACGGTTGATTTCTGCAT
TTCCATCTGAGGTACCAGGTTCATCTCACTATGGAGTGCCAGACAGTGGGCGCACGTCAG
TGGGTGTGCGCACCGTGCGCGAGTGAAGCACGGTGAGGCATTGCCTCACTCACGAAGTGC
AAGAGGTCAGGGAGTTCCCTTTCCTATTCAAAGAAAGGCATGACAGATGGCACCTGGAAA
ATCGGATCACTCCCACCCGAATACTGCGCTTTTCCGACGGGCTTAAAAAATGGTGCACCA

25736 AAATCATGTACATAGATATTATAATTTTTCAAATGCTTGGAAATGTCAAATTAAAATTAT
GGTTGATTGTATTAAATAGATACATATATGATAACATAAAAATATGAAGAAAAAGTAAAT
ACCAAATAAAATGGCTCAAATAATTAGAGATTAGACAATTAATTAGACAATAATTAGATC
AAATACAATCAACTCGAATTTATTTAATTAGAGCAGATGCTAACTTAATCAGCGCCTGAT
TCCTGAGGTAGCAAAAAGTCTAGGTGGAGAGAGAAACTTACCCCTTTTCTTACCCTTCCT
[C,T]
GGTCATCCTGGGAGCTCCACTTTCCTCTGTAGAATTTATTCAGCCTCCTTAGTAAACATG
GACTTGGTCCCAAACAGGTAACCCAACTGACCACAAGAAAAGCAGCCTAGATCCTGAGCA
TTCAGCTCCTGTCTTCACACAACAGACACCACCTCAGTCCCATCAAAGCCTGTGAAGTTT
CCCTACATCCACCATTGAGACATATTCCAGAGCAGCCTCTCAAAATTGCCTTAACAGGAT
GGGACACGATATGGTGGAGCTCCTGGCTCAGGACAGCTGCCTCACCCCTTCCTACTGAGA

26321 CCCCTTCCTACTGAGAAGTCTGTATCTGCTGGTTAGAGCTATCAAACTGTAGAAGGCTTA
GTGCCTGTCCCAGCAAGTGTCCCCTCAAAAGCCTTCTTGTTTTCTTTCCTTCTGAGAAAA
GCATACAAGAATGAGACCTTCTATGTTAGAGAGAACTCAGCCTCCACTCTCAATTGACTT
GGTTGACTGATGAATTGATGCCCTGAGGAGGGGATAGATTCAGGGAAGAGACTGTGCTGA
ATGAGTCTGTGTTTTCCTAGCTTTGCTGTCTGTGCAAATAGTGGAACCCAGAAAAATATC
[G,A]
GGTGGTAGACACACAGACACTCTAATTGTCTGAATTTAAATATTATTTAAATGGAACTTA
TAGTATCATTATATATTGATACCATAATATCACATAAAATTTGTTTGATATATAAACAAA
TATTGATATTTTATCATAATATCATAAAGCAGTTGTGCACACAATAGCAGATAATATTCT
CCGGTTCTATAAAGTTTATATGTTAATGTTCTTACAAAATTTAACTCAGTTATTTTATAA
AATTACCTAACAAAATTTTGTTACTGTGCTATCATAATAATACATAAAACTATGAAATCA

```
31359   ATAATAAAAATCGTAATGTTAAATAATTTCATTTAAAAGTTTGACTAATTAGGCATATAG
        AGAAATGGGCAGTATGTTGACCAGTAAACAGGATACAAATTATTTTTCGAAAAGCAAACA
        ACATTTTTATATTTTATTTATTTATTTATTTTTCTGGTCGGAGGAGCTGCTTTATTGTCT
        GAGGACATAGTACGGTCCTCTCCCTGGGAGGTGGGGTCTTCCACCGGTCACCCGAGGCAG
        TGTTCCAGGAGGCTCCATGCAACTCACTGCTGGGCTCAGCTGGGGGCTGGGCCTTGGAGA
        [A,G]
        GGCGAACTGTGCAGGGAAGCAGTAGCTGTGGGTCCTCACCGCCCGCTCTGCTTTGCTGCA
        CTGGGTCCCTGGTGCTCCTCGAAGTCCCACTGAGCCTGCGTCTCTATAGGACGGTGACCA
        TCCAGCCCAGAATCTTCTAGTCAGAGCACAGTTTGACCAGGCCAGGCATTTCCGCTTCCT
        CTCCTTGGGCTGGACTTTGCACTTGGGTTTCTTCCAGTCCTTCTTCTGCCGCCTGTCTGC
        CAGAGCTTAAATTCCAGCCTCACAAATGTTCCAGCTAGAAAAGGCGTGTCCGCGGCGCCT

35098   GGACAAAGCATGAGATAGGTTTTCTGAAAATTTCCCAGAAACTAAACCTTGAAAAGATGA
        CCTAGGGACTAGGCCTTGGGTAAAAAGGGGCTGTCGAGTGAGGGTTAGTCATAGAAGAAA
        ACTCAAGTTTTTTAAAAAATAGAGTTTGGAAACTCTTATTTATTTTATTTTATTTTTTGC
        AGAACTTTTCTCCCAAAAAGAGTCTGTGGCACAGTTTACCCCTTCCTGATTCAGAAATGT
        GTAATAAAGTTTGGTTTGCAACTTTTCAATGCCATTTTTTTAAACTAATAAATAGTGATT
        [T,G]
        AATCAAGTTATGCAGTAAGTGGACTAAAGTTTACAGGGCACACATGAAGTGTGTCACACT
        TCATTATTTTATCGTGTCGTCTATGACATCCCTGTGAGCACAAAGCCCTCTATGCACAAT
        TCATATTACCACTACTGACGTCAATATACATCTTGTCTCTGTCTCCTCCTTTCCCAGCAN
        GACCCCCAGCAGNAGAAATATTCAAAGTGTTAAAATAAATCTGCTGTATGCATCCT

40532   TCCTAAGAATTACCAATGGGAAAATTGGAGTCTAGAAAATGTTGCCATCATTTTAGCCCA
        CCGGTTCCCCAATAGCTGTATTTAGGTGATAAAGTGCTCTCGAATGCATTGCACAGACTC
        AGTTGCCATGACGATTTTGTGAAAAGTAACATGTTTGGTTCCCCAGAACACAATACTGAC
        TCTGGAGCTTTTAAGCACCTTTGTATTATTAGTTAATGCTTTTAAGTCATAATAGTTTAT
        CAAAGAAAAATTTGAACGGTTGGAATAAGGACTCCACGCATGTAATTGCAGTCCAGTTTC
        [C,G]
        TGATATTACAAATCGTTTCCAGGGAGAAAAAGAGAGGACCTGTGAAAAACCTGATGAGTT
        GGCTATGAGTTTTTATCCACCATCACTAAATGATGCATCTTTTAATTTGACTGGATTCAA
        TAAAGATTGTGTTGTTTTGAATCAGTTTCTTTTTGAATTGAAAGAAGCCAAGAAACGCAA
        AGACATAGATACTTCCATTAAAAGCATAAGAACAATATATTGGCTGGAGGGTGGTCATTC
        CGTAGGAAGTAATACCTGGGTTACTTATCCACAAGTCTTGAAAGAATTTGTACAGTGAGA

41706   AAACTTGAATGGAAACAGTTTATTTCTCTTCAAACAAAAGCAAAGGCACAATGTTGTTTC
        CTATCATTTTGGAATAACTGTACCCTGCCTCTTGTGTTTTGTAAACTCATTCACTCATTC
        TTTAATGTGCCACCAAGTACTTTTTTCTTGAGAGTCAAAATATATTTGTTTCACAATGTC
        CAAAAATGTGCAAGAATGTAAAGCTGGTTTTTAAAAACATAGCCATGTGATGGCATGTGC
        CGTTAGTCCCAGCTACTCAGGAGGCTAAGGCAGAAGGATCCTTTGAGTCCAGGCTATAAC
        [A,G]
        CACCATGATTGTGTTTGTGACTAGCTACTGCACTCCAGCCTAAGCAACATAGTGAGACCT
        CATCTCAAAAAAAAAAAAAAAGAAAAAGAAAAGAAGAAAAAAAAGAAAAACAGAACAAAT
        TAGACTTAAGTACCATCACTTAATTTTTAGTTGACAGTCTTTAGTTGATTGTTTTGGATA
        AGACATTCTGGGGCTTCTTGAATCTTGGCCAAAAACCAGTTGTTTTTGAAAACTGTTTTA
        AATTAAGCATATTTATGTATTTTGGATAAAAATCAACTACAAAGAAAATTTTATTTTTTT

51095   ACTTCAGTGACAAGAGATAATTATCATCTTCTGACAGAGGAGGAATTTGGGGTTTGGTCC
        CAGTCCATGAAGTGGCACAGTCAGAATAAAAGGTGAGAGCTTAGGAGATTAGCGGAGGGT
        AGAAGAACACTCTGTCTTGTGACCAGCTTCAGAGAGCCTGGGGCCATGGCTTCCTGGTCA
        ACATTAGGCCCTGCTGCATGGTGACCCCTGGGCAGGCAGTGGGAAGCCTGAGGTGTGGCT
        CCTGGTGGCCTCACAACTGCCACTCTTTCCTGAAGCTCCTATTTGTTCTGTCAGCTAAGC
        [G,C]
        CCCATCCCAGTAGGCCAGCAACACACTCAAGACCAAGAACAGGCCATGGTGAATCTCAGG
        GCCACTGAGTGCCTGGGCTGGCAGGGGCAGAGTTCCTCAGGGCTCAGTGACATTTGGACT
        GAGCATGGGCTTTGGAGTCATACAGCTACACTGAGCTCCCAGCTTCACCGTGATCAGCCC
        TGTGTCTGGGACAGGGGCCTCACTGTTCTGGAACTTGAGACGCCATAGTCATAAATTTAA
        CACACACTTCTAACTGCTTTTTCTTTTTTATCTGTCTTTCTCTATAATCACCATGTACTA
```

53101 ACCAAGGCCATATTGTGATTTAAGAAACTAGCAGACAGGACTTTCAATTGACTTGATATG
ATTTATTATTTTTACTACTTATAAGAATGGAAATAAGTTCTCCTTAGTTTTTTTCTTGGA
GAAAGTCTGACATGTGAGGCACAGATGAGTTATTAAAGGCAGATGACTTTCCAGCCTTGT
CTTAAATGTTCCATTCTTTACCTTAGAAATTATTTAAATTTGTGTCTTCCAAATACTGTA
GTAATATTGATGCTCCAAAGAGATGTCCCACGGAGATTCTGCTCTTGTGTGTCCACCCTG
[G,C]
AGGGAGCTGAGGCAGTTTCTTATGACAGTTTCAGAAGCGAGTAGTCGTGCAGTACTTAAT
CTAAAAAACTTAATGGAAACATGAATTAAGAGAATGATCACTGTTTAGTTCTATCAGCAA
ACTATTAAAAGTGATCCAAAGGAGGTATTTATAAAGAGATATTAAAAGATTTTTCAAGGG
AGCCTTATTCAGGGCAGAAACGCAGACACTATCGCTGACCTCACCACAGAAAATACCCTC
ATGGGTTGGGAGGGACCAAGGGACGCTCTGGTCCTGCTGACCTGCATTAATCACAGCCAG

54556 GAAAGGGAGGAGCGAGAGAGGGTGGGGCTGAGGAGGAGGCTGTGTGAGAAAAGGGGCGGG
GCGCGCCCAAAGTCTGGCGGAAGAGCGTTACCCTGGCAACCCTCCCGCGGAGGCCGAGAG
AGGCCACCGGCCCTTTGTTGATCTGCGAGAAATCAAACTACGAACACGACAAGCATTAGC
CTGCAGCTCGAGGAGACAAGGTGTCACAATTACAAGGGGAAACTAGCCGCCCTAGCTCCA
CTGTCTCCCCAGCACGAGAGATTTGAGAACAGAAAGGCTTCCCTCCGCAGGGCGAAACTG
[G,C]
TGGGCTGCCTGGAAGGGCGAGGCAGGGAGCGGAACCGTCTTCAGGAAATTTCGGGAGTTC
CGGGGTCAGGTCCACTCCCCGGCTGTTGTTGTGTTGTTGGCAGGGCAGAGGGTCTAGGAT
GCCAGCCTGCTCCGGGCTGCGCTGTGCGCCTATCCCAGGGCGGGGGGATGCGGGGCGACA
CCCGCCTCCCGGTGCATCCAGGAGTTGTAGTCTCTTCACCGGTTCCCCACTGTGGGTGGT
GGGGCTGCAGGAGGACAATTCAAATTGAGATAGGAGCGGAGGCGGAGCGCGGCCGTGCAG

61872 ACACTTCACATCAGGGTTCCCAAATGTATAAAGTTAACATTGACACAAATGAAGAAGGAA
ATGGCTATGCAAAAATAGTAAGAGACATAATTACCCCACTACCCACTATCAGTAATGAAT
AATAAAGCCAGACAGAAAGTTAACTTGAGAACAGAGAATTTGAATAACACTGCAAACTCT
AAACCTAACAGACATATAGAAAACACTAGTCACAGTGAATAGAAGTGAAAAAACAAAAGA
AACAAACACTCCACACAGCAATATCAGAATATACAATTTTTTCAATAGGTCATGAAACAT
[-,T]
CTCCTGGGTTGATGACCTACTAGGACACAAAACAAGTTTTGCTAAATTTTAAAATGGTAA
AATATGGGCCAGGGATGGTGGCTCATGTCTATCATTCCAGCATGTTGGGAGGCTGAATTG
GGAGGATTGAGTGAGTTTAGGAGTTCATGGCCAGCCTGGGCAACATAAGGAGACCTTGTC
TTTACAAAATATAAAATTAAAAAATTAACTGGGCATGATTACACGTGCCTGTGTGTCCAG
CCACTCAGCAGGCTGAGGTGGGAGGATTGCTTGAGCCTGGGAGATCAAGGCTGTGGTTAG

62172 TCTCCTGGGTTGATGACCTACTAGGACACAAAACAAGTTTTGCTAAATTTTAAAATGGTA
AAATATGGGCCAGGGATGGTGGCTCATGTCTATCATTCCAGCATGTTGGGAGGCTGAATT
GGGAGGATTGAGTGAGTTTAGGAGTTCATGGCCAGCCTGGGCAACATAAGGAGACCTTGT
CTTTACAAAATATAAAATTAAAAAATTAACTGGGCATGATTACACGTGCCTGTGTGTCCA
GCCACTCAGCAGGCTGAGGTGGGAGGATTGCTTGAGCCTGGGAGATCAAGGCTGTGGTTA
[G,C]
CCATAATTGAGTCACTGTGCTTCAGCCTGAGTAACATAGCAAATCTCTGTCCCAAAAGAG
ATTAAAATATTACAAACTATCATTTTTGATTAAAAGGGAATACAATGAGAAATCAATAGC
AGAAAAAATACTGGAAAATCTACAAATATGTGGAAATTAAACAACCCACTCTTCAGCATG
CTCTCGTTAAGGGTCGGAAGACAATATTGTGAAGATGTTCACACTACCCAAAATTATCTA
CAGATTCAATGTGATCGCTGTCAAATTTTAACTGTCATTTCATTTGCAGAAATACAAAAA

62860 TGACTTCCTAATGTATTACAAAACTACAGTAACCAAACTATTTGGTACTGACATAAAGGC
AGACAGACAGACCAATGGAACAGAATAGATCACAGGAATAAACTGTCATATATATGGCCA
AATGAGGAGTTATTTTTATATCCATATTCATTGCAGCATTATTCACAACAGCTGATAGGT
GGAAGGAACCCAAATGTCCCTCAGTGAATGAGTGGATAAAGGCAATTTGGAATATACAAA
TAATGGAATATTATTCAGTTTTTTAAAAGCAGGAGATCTGATTATTTTTACACTAAGAAT
[A,C]
AATCTTGAGGACATTATGTAAATGAAATAAACCAGTCACAAAAGGACAGACACTGTTTGA
CTCCAGTTAAATAAAATATCTAATGTAGTTAAACTCTTAGAAACAGAAAGTAGAATAGTA
TCAGTCAGAGCCTTAGGGGTGGAAATAAAAGGGTAGTTGTTGTTTCATAGGTATTGAATT
TTAGTTTTACAACATAAAATCATTTTAGCGATATGTTGCATAGCAATGTGAATATATTTA
ATATTATTTAACTATGTACTTAATATATTTAAGATGGTACATTTTATGTGTTTTGAGTAC

67086 AGGATATATGGGATTAATTTTGGGTAAAATCCATCTCAACTTACAGGACATTACTGTAGT
CCCAGGAGTTGTTGACTGTGATTATGAAGGAGAAATTCAAGTAGTGGTAATATCACAAGA
TTTGTTAGTTTTTGAACCTGGAGAATATGTAACTCAACTACTGCTTATTCCCTGGGAGTT
GTTTCCTTCTCCACATAAGGAGAAATGAGAGAATCAAGGATTTGGGAGTACAGCTAGGAG
GAAAATTTATTTATCACAACCCATAGCATCTAATAGACCCACCTGTACAGTGCAAATTAA
[C,A]
GGAAAAAAATTTCTATGGGCTTATGGATATGGGAGCTGATGTGTCAGTAATATCTAAAAA
CAATTGGCCCCCATCCTGGCCCCTGCAATTAACTCCTACATCGCTAGTGGGAATAGGAAC
AGCTCAAAGTGTTCAACAGAGTGCTGAAATTTTACCCTGTCTCAAACCAGATGGACAGTC
ATGTACTTTTAAAATTTATTTTGCAAATGTAACTGTTAACCTATGGGGCCAAGATTTACT
TACAGCATGGGATATAAGACTTGCAAATGAAACTATTGACAATCCAGGGTTCAAAATGTT

67621 TTTACTTACAGCATGGGATATAAGACTTGCAAATGAAACTATTGACAATCCAGGGTTCAA
AATGTTAAAGAAAATGGGATGTCAGGCAGAAAAGGCTTAGAAAAGTCCCTACAGGGAAAC
ACTGATCCTATATCAATAGCTGGGCAAACAGATAGAAAAGGGCTAGGTCATCAGAATTTC
TGATGGGAGTCACTGATATTTCTCCCCCATCTACTGTTTTACCGCTGGAGTGGCTGACTA
AAAAACCTGTATGGGTGGATCAGTGGCCCCTATCACAGGAGAAACTAACACAATTCCATC
[A,C]
GCTAGTAAAAGAGCAAATGGATGCAGGACATATTGAAGAGTCAGTTAGCACCTGGAATTC
ATCAGTATTTGTAATTCCTAAAAAGTCAGGAAAATGATGACTGCTACATGATTTGAGAGC
TATTAATGCACACATTAAACCAATGGGTGCATTACAGCAAGGTCTGCCATCCCCGGCAGC
CTTTCCAGGAGGCTGGCCTCTCAAAGTAATATATCTTAAAGATTTTTTATTTATTTTTTT
ATTTTACTGTTACATGAGCAGGATAAGCATCGATTTGCCTTTTATGTGCTTTCTGTTAAT

70582 TGTAATATCCTGTGCAGTATGATTTTTCTGTGCAGAAGCAAAAACATATTGGGCATATTT
TCCTAACCCACCGGTAGTGTGATCATACTCTGAAGCAGCACTCCTCCTGAGATATATCAT
GATCAAGGAGCATCAGTACCAGGACCTCTAACTCCCCCTGACACAGAGCAATTAGACTCT
CATAACAATGGTATCAATTATACCACTCCATTGGAGGGACTTCCTTTATGTGTCACCCAG
GATACATTGCTCAACTGCAGTTGCCTTGCAGTTTGATCCCAAGCATGGTTGAGTTACCAT
[A,T]
AAAAAATTATGTACCTATTAGACCTTAGCTTTATTAATATTACTTGTGTAGTTACTAATC
ACTCCTGGCCCCATCACCCAAATTGTACTGATTATACAGAATGGGCTCCCTTTGATAATT
CTCACCCCCCTCCTTGGGCCCACTGTCTTGGCCCCTTAGCTAGACAATAGTCCATGTTAA
TGGGAGACATTATTGACTGGGGTCCCTGTGGTCATTAAGATGGGAGAGATGAGAATCAGA
CCACATGGCATAAACTTCACTGGCACTGGTGGCGAAACTTTAACATCTCTTCACTTCAAC

74175 GCTGCGCCTGCAGCCCCCACTGAGCGTCGGACTCCTTCCTGGAGTAGGGAGGTCCTGTTC
CTTCTGGAGCGACAGACACCCTTTCTCCTGGCCTTCTCGCTTACTAGCCCGGCAGGTGCT
GGACAGGAGATCTGAGCTGGTCCTGCGTCTCTGAGGAGCTAGGAGCCCGGCTGGGAGAAC
AAGGAGACGAACTGTGGGGAGAAGGGGCGACAGGAACGCCAGGCTCATGGGACCGCTGGC
AGCGGCCTGGGTATGGCTGGCGGCTGAATGGTCAGAGATACGAGAGGTGGCCACTGTCCC
[C,-]
ACCTTTGGCCCCCTAGCCGGCATTCGTACATTCTGTGCTCAACAAACGGAAGCGGCAGCT
GGAGCTGCTGCTCCGGGAGGTGGAGTGGCCTGGCAGAGGGCACATGGCTGCCACCTGCTG
CAAGGTGAGCTGGTCTGCAGCCTGGGCCCACAAAAGGCCGCTCTTGTGCAGGACACACCG
CTGCCCTTGACCCTCTTGCTCCCCCGCCTGCTGTGCAAAATGCTCAGGTCCCTGATCTCG
GGCTTTCCTGGCAAGTGCACTGTGGTGGGGAGGCAGCAGGGAGGAGGGCTTTTCCAGGAG

74478 CTTTGGCCCCCTAGCCGGCATTCGTACATTCTGTGCTCAACAAACGGAAGCGGCAGCTGG
AGCTGCTGCTCCGGGAGGTGGAGTGGCCTGGCAGAGGGCACATGGCTGCCACCTGCTGCA
AGGTGAGCTGGTCTGCAGCCTGGGCCCACAAAAGGCCGCTCTTGTGCAGGACACACCGCT
GCCCTTGACCCTCTTGCTCCCCCGCCTGCTGTGCAAAATGCTCAGGTCCCTGATCTCGGG
CTTTCCTGGCAAGTGCACTGTGGTGGGGAGGCAGCAGGGAGGAGGGCTTTTCCAGGAGCC
[T,C]
TGAACAGAGGATCTTGGCATAAAGAGGAGAGAGAGGTGGCTGACTGGTTCCACTTGTAGG
TAGGGGGGCAACAAACCCCATGGGACCCTGTTTTTTCAGGGAGATTTCAGTTCACTTCTT
ATCTTTTCTCCACCCACTTGAGCCTCTGAGAATAGAGGAGACGAGGCTGTTTTAAATTGG
CCTAACCATAATGGTCTGGACCCTTGCCCCAGGGCAGACCTAATTTTGGGGCTCTTTGCA
GCATGGAGGCTCACGCCTGTCCACCCCAGGTGTCTTCAATATAGGGTCTAGTTAGGCCTG

77092 TGCAGAATGCAGATGGATCCAAAAAATCACAGGATGTGGGAGGTGAAGGAAGAGCTTGTA
AAAATGAAAAGTGGCTGGTGAAAGAGTAGGAGGCATGAAGAGGATGAGACCTGCCTGGGG
CAGTGCACATGTTTTGTTCCAGCCAAACAATCAGATGAGGTCTTGGTCTTGGACCTGGTG
CCAGGGAATTCATAAGCCCCCTTTTGCTGTGGCCTGGGAGCTGAGGTCTTTGGTTCTGAA
ACCAAATGTAAATTTTGGACTCTGGAATACCTGTCTGTTTAACCAGTTTCTCTCTACTGG
[T,G]
GATCGCAGGAAATGAATTATCCTGTAAAATTTTGTGGATTCTTCTCAAAGGCTTCAATGA
GTACACTGATTTGTGGTTCAGTGATGGATGTCCCTTTCCTCCTGCCTTCTTATTTGACTT
ACACCATCAATATTAACTATGGCAGTTATGGTAATATCATTCTTTACAACAGAGGAAACT
TCAAGTCATTCATTGATGATATCAAAGCCTCATTCTCCTACATTAAACATTCTCCTACAT
TTAGCTTCTTGAATCTTCTATGTCCACTGTCTAGAAAACCTAAACACATGAAGTACCACA

77328 TGAAACCAAATGTAAATTTTGGACTCTGGAATACCTGTCTGTTTAACCAGTTTCTCTCTA
CTGGTGATCGCAGGAAATGAATTATCCTGTAAAATTTTGTGGATTCTTCTCAAAGGCTTC
AATGAGTACACTGATTTGTGGTTCAGTGATGGATGTCCCTTTCCTCCTGCCTTCTTATTT
GACTTACACCATCAATATTAACTATGGCAGTTATGGTAATATCATTCTTTACAACAGAGG
AAACTTCAAGTCATTCATTGATGATATCAAAGCCTCATTCTCCTACATTAAACATTCTCC
[T,C]
ACATTTAGCTTCTTGAATCTTCTATGTCCACTGTCTAGAAAACCTAAACACATGAAGTAC
CACAAACTAGAAATAAATACCTCAATAGGCACCAATCATAAAAGTAAATCCAAGAAGGAA
CAGAATATATGAATACATATATAACAAGTAAAGGGATTCAATTAATTAAAAATCATCACA
CACAAAAAAGCCCAGAGTCATGTGGCTTAACTGATAAATTCTACTAAACATTTAGTGAAG
AATTAATGCCAACTCTTCACAAGGCCTTCCAGAAAATAGAAGACAGTTATTGGGAACACT

77385 CTACTGGTGATCGCAGGAAATGAATTATCCTGTAAAATTTTGTGGATTCTTCTCAAAGGC
TTCAATGAGTACACTGATTTGTGGTTCAGTGATGGATGTCCCTTTCCTCCTGCCTTCTTA
TTTGACTTACACCATCAATATTAACTATGGCAGTTATGGTAATATCATTCTTTACAACAG
AGGAAACTTCAAGTCATTCATTGATGATATCAAAGCCTCATTCTCCTACATTAAACATTC
TCCTACATTTAGCTTCTTGAATCTTCTATGTCCACTGTCTAGAAAACCTAAACACATGAA
[G,A]
TACCACAAACTAGAAATAAATACCTCAATAGGCACCAATCATAAAAGTAAATCCAAGAAG
GAACAGAATATATGAATACATATATAACAAGTAAAGGGATTCAATTAATTAAAAATCATC
ACACACAAAAAAGCCCAGAGTCATGTGGCTTAACTGATAAATTCTACTAAACATTTAGTG
AAGAATTAATGCCAACTCTTCACAAGGCCTTCCAGAAAATAGAAGACAGTTATTGGGAAC
ACTTCCCAATTTGTTCTATCAGGCCAGTATTACCCTGATACTAAAGCCAGACAAAAGCAT

77947 GGCCAGTATTACCCTGATACTAAAGCCAGACAAAAGCATCACAAGTAAATATGAACATAG
ATGAATTTCCCTGATAAATACACAAACAGAAAATCTCAAAAAAGAGTGAAATGAATCAAG
AATAAATCAAAATGACTGTACACCATGACCAAATGGAATTATTTCACAAATGCAATATTG
ATCTATCCAATAATCAATCAATGCATTACACAAAGTAATAGGATAAAGGAAATTAACAGA
AAGGTCCTTTCAACAGACACAGGGAGCATTTAACCAATCCAATATTCATTCACGATCTCC
[C,T]
TGGAAAAGAGGAGTACAAGAAACTTCCTAGATCTGCTAAAAGGCGTCAATGTAAAACTTA
CAGCTAACCCCATAATAATAAAATACTGGTTGTTGTGTCTTGACATTTGAGAACAAGACA
AAAATGTTAACATCCAAATAAATTACATAAGAAAAATAAACAAAATCATCAATGTGGGAA
AATAAGAGGTTAGAACTCTCTATCATTGCAGAGGACATAATGTGAATATAAAAGTTTATA
AGAAATTCATTAAAACTCACTACAACCAATAAGTGAGTTCAGCAACATCACAAGATACAA

79395 CTGAGCATGGTGACATAGGGCATGCCATTGACACCCAATTTGCTGGTGTTCACCTCCACA
TACTAAGTTGCTAGAGAATTTGCAGTGGATTTCCATTTTGCTGCATCTGGCTGTCCAGAA
GCCCTGCAAGTAGAATGGGATGGTCAGGAGAAAACATTGAAAGAATAGATGGGAGTTCAG
AGGCTGCCCACCCTCCTGCCCTCTGCCCACGGGCCACAGCCCTCACCCAGCTGTCCAGTG
TGTATGTCTGCTAAAGGCTGCTGCACTTGTTCTCCATCACAGAGGTGGGGTGACAGCAGT
[G,C]
TGTGGGCACCACACTCCATGATCAGCCTCTACTGTTGGTGCCACAGCTCCAGGTAGAAGG
GCAGGTGAGCCAAAGACAGGTCCCACCTTCCACATCCAGCCCACATTCCCACCAACTTCC
AGGCCCACCTCCATATGCTGATGTGATGCTCTCCTTAGAGCTCTTGTGGTTCTTCAGCCA
GGAGATGGAGAGAGTGGGGTTTCAAGAGTAAGGCAGTGAAAGTTGGTTTGGGTGGCCACC
ATGGTTGGTGGTTTCCTGTCCATCCACTCGGTCCAAGTCCAGTAAGGGGCCTCTGCTGGT

81111 ACTAGGGTGTGTGGGCAGGGGCAGAGGCCATTGAGGCAGGTGAGGAGAAATTTTCATCCT
CTTCCTGGTCTGCCCCTCTCCTGGGGTCTAATTTCCTCTATTCCGTCTGCCTCTGGCTCC
CTGGCTGGCTCCTCTTCACTCTCTTGCCTGCTCACCCCAGAGGTCCCAGGGGCTCAGCCC
ACCACAAATGGTCCCCAAGTTGTAGCTGACCCTTCCATGTCCATCCCATGAGGACCCTCA
TCTGCCTGAGTATATCTCTGGGCTCCTCTGAAACCAGAAGTCCCACCTCACTGACTGCTC
[-,C]
ATGGCTAGGCAGCATCCACCTGCCACTGTTCCAGGCCAGAATGACTGGGCATTGTCCCCC
TGCCTGCTCCCTCACACCTACAGCTCATGCCCCCAAAATGCTGTTGGCATCCTTCATACA
ACCCTCACAGCCACCCTGCCCCCTGCTGGGCTGTAGGTTTGGTCTCCTGGTGCCTTAACC
CCTTTGTCCATCTGCCCCTGGGCAGCCACCTGAGCCCCTGAGCACTGCTGTGCTCACATG
TGCAGTAGCCCCCTCACCCAGAGCCAGCATCGAAGTCTCCACAGGCCAATTCTGGCCTCA

81610 TGGGCAGCCACCTGAGCCCCTGAGCACTGCTGTGCTCACATGTGCAGTAGCCCCCTCACC
CAGAGCCAGCATCGAAGTCTCCACAGGCCAATTCTGGCCTCATCACTGCTCCTGGAACCC
CAGGGCCCTGTGCCCCAATCTCCCCATCTGCAGCATGGGTGTCTCTTCCTACCCCAAGCC
TGCCCCCCAGAGCTCAAGACATCCAGAGCCATCTAATACATATGTAATACATACAAATTA
CATAACAATTTGTAATATGTTGTACTACATACACATTTTCATATGAATTCATCATAATAC
[G,A]
TACAAATTATGATGTCATAATATATTGTGATGTGACAATACACATGAATTATCATGTCAT
AATACATTGTGATGTCATAACACATACTAATTATGATGTCATGATATATTGTGATGTTAT
AATGCATATGAATTATGGTGTCACAATACATATGAATTATGATGTCATGATACATTGTGA
CGTAATAAGAATTGTGACATCATAATATATCATGATGTCATATGCATGCAACTTATGATG
TCATGATATACTGCAATGCCTTAATACAAACCAATTATGATACAGTAATATGTTGTGATG

Map data:
chromosome 16